

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



GREEN LIGHT FOR TRANSPLANTATION

Enhancement of stem-cell engraftment in a zebrafish model **PAGE 468**

AGRICULTURE

FARMING TOMORROW

Crop production in a changing climate

PAGE 396

BIODIVERSITY

ENGAGE, DON'T RENEGE

Police offsets, and monitor from space

PAGES 401 & 403

NANOMATERIALS

COMPLEX DNA STRUCTURES

Graph theory paves the way for 3D nanoscale printing

PAGES 412 & 441

NATURE.COM/NATURE

23 July 2015 £10

Vol. 523, No. 7561



THIS WEEK

EDITORIALS

REPLICATION Rack up the phone bill for good research **p.382**

WORLD VIEW Russia increases pressure on science and society **p.383**

THUMBS DOWN Proportion study undermines theory of our special hands **p.385**



Prepare farms for the future

Scientists must work closely with farmers to ensure that agriculture can stand up to the ravages of climate change.

Ignore the climate sceptics who set up a straw man of the need for 'settled science' and then burn it to the ground. Ambiguity is the acknowledged refrain of the climate-change symphony. From storms to sea-level rise, all projections of future change are surrounded by a residual uncertainty that will not go away, no matter how sophisticated our climate (and climate-impact) models may become.

The future of global agriculture is one of the most urgent issues in a warming world. Farmers must prepare for, and adapt to, a changed climate that is likely to feature more erratic rainfall, temperature extremes, drought, soil erosion, invasive weeds and durable pests. Science, error bars included, has much to offer these efforts. But if adaptation is to work, climate scientists, agricultural researchers, farmers and government officials must work closely together.

As a reminder of how sensitive farming is to extremes, consider the record-breaking 2003 European heatwave, which caused more than €13 billion (US\$14 billion) in damage to agriculture and forests. In less-developed parts of the world, prolonged drought and other extremes come with even more direct social costs, in the guise of increased hunger and risk of violent unrest.

Reliable climate services, such as those being established around the globe under the auspices of the World Meteorological Organization, can provide valuable early seasonal forecasts to farmers and governments. Their accuracy and coverage must improve in the face of the coming climate crisis. But the strategic decision-making that climate change will increasingly force on the farming sector requires forecasts that look further ahead. And climate change is far from the only uncertain outcome that farmers must grapple with as they prepare for the future. Trade, technology and socio-economic change can affect agriculture just as profoundly as changes in rainfall and temperature.

Farmers are natural adaptors. They have been tweaking and changing their practices since humans first began to grow food, and most today have a keen sense of what works best on their fields. But climate change may require drastic measures beyond the capability of individual farmers, from expensive irrigation schemes to the transformation of farming systems. These may not materialize through economic growth alone. And specific needs and adaptation options will substantially differ from region to region — or perhaps from village to village — depending on farm types, soils, local climate and topography. There are as many different ways for agriculture to adapt to climate change as there are different types of agriculture.

Models of different scenarios concerning crops, climate and economics can help, but only up to a point. Agriculture is an early adopter when it comes to using science to inform and guide adaptation. However, this use of science does not rely only on the scale of models and the skills of modellers: trust, intuition and cultural empathy are just as important.

Developing an improved crop variety in the lab is a very different

thing from convincing farmers to adopt conservation agriculture, switch to semi-arid farming systems or do anything else that may not come with an immediate, tangible benefit. To produce any 'actionable' outcomes, the science of climate-change adaptation must therefore engage and listen to the people it is supposed to serve.

As we discuss in a News Feature on page 396, adaptation researchers are increasingly aware of this communication challenge. Science-led

"The science of climate-change adaptation must engage and listen to the people it is supposed to serve."

initiatives, such as Modelling European Agriculture with Climate Change for Food Security and the Agricultural Model Inter-comparison and Improvement Project (AgMIP), are being pursued in close consultation with local experts and farming communities. Such programmes are a valuable step beyond coarse academic projections of climate impacts such as changes in global

crop yields, which lack regional specificity.

Regional studies suffer from the inevitable uncertainty over the magnitude and manifestations of climate change, and perhaps even more over the course of socio-economic and technological development. But carefully crafted regional case studies, informed by locally sourced data, can produce plausible future scenarios from which local planners can draw a range of tailored adaptation options.

AgMIP aims to produce a standard experimental protocol to study climate impacts on farming, which will help adaptation efforts even further. If it succeeds, the programme should solidify adaptation research, in the same way that model comparisons have improved the consistency of the physical climate sciences. The future is uncertain, but that cannot be used as an excuse to fail to plan for it. ■

Timeless advice

The best guidance on how to get ahead in science stands the test of time.

How can a young researcher get ahead in science? They need perseverance: "You do experiments and 90% of them aren't going to work. Nobody warned me about that." Boldness: "People don't ask enough questions. They're embarrassed." Mastery of the basics: "I didn't even know where the pipettes were." And perhaps a dose of reality: "Rejection is an ever present companion in science."

Those quotes, all from researchers interviewed for a Careers Feature on page 491, demonstrate that there is more to a successful scientific

career than being good at science. And although opportunities for paid positions in research have flourished in recent years, so has the competition. The message has yet to filter down to schools and university undergraduates, but professional science has become one of those careers that teachers and lecturers could euphemistically describe as ‘popular’ and ‘competitive’.

This is good for science overall. The global talent pool is well-stocked and the number of proficiently trained apprentices eager to take their chances is healthy. It is less promising for the scientists themselves: too many are chasing too few positions.

In such a climate, providing careers advice for scientists has become a career in itself. Yet, as the researchers highlighted in the feature make clear, many of the questions and anxieties that trouble early-career scientists also crop up in other careers. And the useful skills that ambitious researchers are urged to develop are hardly unique to science either: confidence, communication skills, networking abilities and persistence will help to propel people up the ranks in most professional fields.

Not everyone is suited to a career in science — nor is there space for them. So how can the community identify and help those young researchers who have the best chances of success? Senior and established scientists can help through formal mechanisms such as mentoring schemes and more informal routes, including workshops and blogs. Universities and other institutions should recognize that these contributions are valuable, and assess and reward them appropriately.

Amid all this advice, how should young scientists judge which guidance to listen to? *Nature*’s advice to these young scientists is to read *Advice to a Young Scientist* by Peter Medawar (Harper and Row, 1979), which celebrates its 36th birthday this year. Back when it was published, digital science meant little more than measuring fingers, and to modern readers the book may look as if it belongs to another age,

but almost all of its content remains startlingly relevant. Furthermore, it is warm, witty and written in a welcoming way that, at the very least, shows scientists that scientists can (a) communicate and (b) do so as well as anybody else.

Here is Medawar, for example, demolishing the platitude that science is based on mere curiosity. “Curiosity is a nursery word,” he writes. “Most able scientists I know have something for which ‘exploratory impulsion’ is not too grand a description ... A strong sense of unease and dissatisfaction always goes with lack of comprehension.”

“How can the community identify and help those young researchers who have the best chances of success?”

But he is not always correct. On scientists who find that the job is not for them and opt out of research, Medawar claims that “the qualifications required of scientists are so specialized and time-consuming that they do not qualify him to take up any other occupation”.

In fact, as *Nature* has argued before, a solid grounding in science and the skills of research offer a strong platform for many alternative careers.

Lest anyone jump on the “him” in the above sentence and assume that this is a book ‘of its time’ that paints a male-dominated picture of science, Medawar is frequently at pains to stress the benefits of and the need for greater equality — for better and for worse. “Men or women who go to the extreme length of marrying scientists should be clearly aware beforehand, instead of learning the hard way, that their spouses are in the grip of a powerful obsession that is likely to take the first place in their lives.”

And on the original point, on how young scientists can get ahead, he writes: “A novice must stick it out until he discovers whether the rewards and compensations of a scientific life are for him commensurate with the disappointments and the toil.” Indeed. ■

It’s good to talk

Help for those struggling to reproduce results could be just a phone call away.

Survey results released last week by the American Society for Cell Biology (ASCB) included an interesting nugget. Some 72% of respondents said that they had been unable to replicate a published experimental result. Yet a higher proportion (77%) said that they had never been told that their work could not be replicated.

There could be many reasons for the difference. The most obvious would be that no one actually tried to replicate the research in question (or that they did not try very hard). When survey participants were asked how they responded to such problems, 55% said that they did not bother resolving the replication issue because they did not think the research was important enough to pursue. For others, the survey results suggest that if and when they did try to replicate, and failed, then they also failed to flag the problem with the original researchers. And it means that they did not ask the people who are best placed to help answer the most obvious question: what am I doing differently to you?

That is not always easy, but it should be the first response. And those on the receiving end of such enquiries should be open to them, not defensive or hostile. As this journal has pointed out before, there is often an art to science. The methods sections of papers, as rigorous as authors and journals try to make them, do not always tell the full story. They cannot pass on tacit knowledge — just as someone cannot be taught adequately from a book how to ride a bicycle.

More than 800 of the ASCB’s 9,000 or so members answered the survey. They reported that the most common way to resolve

problems with failed replication attempts was through collegial consultation with the lab that did the original experiments. In an era of huge competition in biomedicine — when some researchers might fear hostility or even retaliation from senior colleagues when questioning the reproducibility of their work — the survey shows that amicable collaborations, including reagent sharing and open communication, can improve science and make the work of scientists more efficient.

The ‘replication crisis’ in science, and in biological research in particular, is a serious and complex problem that will not be solved by better communication alone. This journal and others have launched initiatives that aim to address many suggested and suspected problems in reproducing results. The ASCB survey results again highlighted some of the issues: respondents rated the push to publish in high-profile journals and poor methodological training as the biggest factors.

The ASCB published a report alongside the survey results, which made some further recommendations for change (see go.nature.com/uh1wsu). These include improvements in statistics training and standardizing the way that experiments are performed.

Even if systemic problems are tackled successfully, some problems of irreproducibility will remain. Biological systems are complex and finicky, and there will always be new experiments, equipment and techniques that take time to master. That one scientist cannot repeat the work of a second does not mean that the first is unskilled or the second sloppy. Although much of the broader media attention on the replication crisis focuses on deliberate misrepresentation and research fraud, scientists and journals

know that the reality is more complex, and less nefarious. Good science is often difficult science. And good scientists should not make it more difficult than it needs to be. So ask for help — pick up the phone. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunqv



Russia's crackdowns are jeopardizing its science

The escalating encroachment on democratic freedoms undermines the nation's claim of support for science, says Fyodor Kondrashov.

For more than a decade, the private Dynasty Foundation has supported science and education in Russia, funding scholarships and organizing summer schools. Yet roughly two months ago, the Russian government applied a controversial law and labelled the foundation a foreign agent. Earlier this month, Dynasty's founder Dmitry Zimin, a physicist turned entrepreneur, was forced to announce its closure. The government's treatment of Dynasty Foundation marks an unwelcome return to the inseparability of science and ideology in Russia.

There is more to these events than a science funder caught up in unfortunate circumstances. There has been a profound political change in Russia, and the causes and consequences of this — for science and for society — need to be examined in historical, political and social contexts.

Reacting to political protests against voter fraud in the 2011 parliamentary elections, the government introduced a series of laws and measures that were designed to restrict foreign influence, but in fact seriously curtailed political and civil liberties. These laws reflect the anti-Western rhetoric of government officials and a renewed popular nationalist sentiment, which intensified last year with the annexation of Crimea and the war in Ukraine.

The law that claimed the foundation was designed to curtail the influence of foreign-funded non-governmental organizations in Russian politics. The fact that Zimin chooses to bank abroad and that Dynasty funded some activities that the government said had the potential to influence public opinion were enough for the Ministry of Justice to target the foundation. Hours after the designation as a foreign agent, Zimin's Facebook account was hacked. Any doubts that the move was political were removed by a scandalously biased report on one of the main government-owned television channels that claimed Dynasty was funding efforts to destroy Russia.

Zimin comes from the generation of my grandparents, with first-hand experience of the brutal application of anti-Western ideology to science. For 30 years, the then-Soviet government deemed genetics ideologically criminal, and students and professors were labelled saboteur agents of foreign governments. In his influential book *Heroes and Villains of Russian Science* (Edwin Mellen, 2000), my grandfather describes the repression — and sometimes murder — of geneticists that forced him, a young biologist at the time, to study the subject in secret for fear of arrest. The ban on the subject led to a collapse of Soviet agriculture. It also caused the heroes and villains of that generation to be defined by their ability to withstand political ideology or to resist compromise with the

regime, perhaps as much as by their actual contribution to science. The generation of my parents was affected to a lesser degree, but they, too, have told stories of withstanding ideology when conducting research. It seems that the Russian government has not learned the lesson of its predecessors, and is determined to interfere with science and use it for ideological purposes.

There are lessons here for other nations and institutions. It is not just governments that use science for ideological purposes. Indeed, scientists and institutions seem oblivious to the moral hazards of mixing the two, and want to consider the ethics of such decisions only in hindsight.

Under the relatively liberal Russian president Dmitry Medvedev, the Massachusetts Institute of Technology (MIT) in Cambridge scored a lucrative contract to help create the Skolkovo Institute of Science and

Technology (Skoltech) in Moscow in 2011. As the behaviour of the present Russian government becomes more totalitarian and hostile to academic freedom, officials at MIT are surely presented with a dilemma on whether to discontinue the collaboration. Institutions and individuals seeking to establish research and academic centres in the Middle East or China have to make similar decisions.

Skoltech is one of several ways in which the Russian government is seeking to promote a pro-science and innovation agenda. A successful research programme must cultivate local talent and attract foreign scientists. To place political ideology that is based on vehement xenophobic rhetoric centre stage in

dealing with a science organization such as Dynasty jeopardizes both. The guarantee of political and civil liberties is an essential condition for the maintenance of a successful research culture, and the ongoing encroachment on democratic freedoms in Russia reduces its appeal as a place for research even further.

If history is any indication, other practices will make a comeback, such as government control over publication or the requirement of political loyalty to obtain funding.

My own generation of scientists will now consider science in Russia not from the perspective of opportunity but from its understanding of right and wrong. History is writing a new edition about the heroes and villains of science around the world. Scientists and academic institutions must remember that the choices they make about becoming involved in projects that are influenced by morally corrupt political ideology will help to determine how history remembers them. ■

Fyodor Kondrashov is a Catalan Institute for Research and Advanced Studies (ICREA) research professor at the Centre for Genomic Regulation in Barcelona, Spain.
e-mail: fyodor.kondrashov@crg.eu

IT SEEMS THAT THE
RUSSIAN
GOVERNMENT HAS
NOT LEARNED
THE LESSONS OF ITS
PREDECESSORS.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/ytrbj2

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

ANTHROPOLOGY

Earliest signs of chicken husbandry

Humans first used chickens for economic gain roughly 2,300 years ago in the Middle East, before Europeans began exploiting the bird.

The chicken (*Gallus gallus domesticus*) was first domesticated in southeast Asia, but its dispersal from that region has been unclear. Lee Perry-Gal and her colleagues at the University of Haifa in Israel analysed animal bones at a site in southern Israel and found a large number of chicken bones, some of which bore butchery marks. Bones from female birds outnumbered those from males two to one, and some showed signs of being from egg-laying hens. The team also saw a large increase in the frequency of chicken bones from the same time period at more than 200 other sites across the region.

Chickens were exploited in this region at least 100 years before they were used by Europeans, the authors say.

Proc. Natl Acad. Sci. USA
<http://dx.doi.org/10.1073/pnas.1504236112> (2015)

STEM CELLS

Heart cells come of age

Human stem cells have been coaxed into forming heart progenitor cells that then develop into more-specialized heart cells.

Researchers have struggled to turn stem cells into large pools of cardiac cells that would further divide. Christine Mummery at Leiden University Medical Center in the Netherlands and her colleagues introduced into human stem cells a version

of the *MYC* gene that they could control. By turning the gene on at key points during the cells' development, the researchers could keep the cells at a certain stage, and expand their number. With further regulation of certain biochemical signalling pathways, the team converted those cells into pacemaker or ventricular cells.

This approach could be used to create new models of human cardiac disease, the authors say.

Nature Biotechnol. <http://dx.doi.org/10.1038/nbt.3271> (2015)



ENERGY

Sun's heat could cut fossil-fuel use

Integrating solar technologies into coal-fired power plants could ease the transition from fossil fuels to renewable energy sources.

Vishwanath Haily Dalvi of the Institute of Chemical Technology in Mumbai, India, and his colleagues looked at solar thermal technology, which collects the Sun's energy as heat. The team reports that injecting this heat into the conventional power-generation process reduces the amount of fossil fuel that needs to be burned in power plants by up to 50%. Solar-aided plants

such as ones in Egypt (pictured) and Algeria are therefore a more economical way of reducing fossil-fuel use than retrofitting existing plants with carbon-capture technology, the authors say.

Widespread deployment of such power plants will require economic incentives similar to those offered by some countries for generating electricity completely from solar thermal plants, they say.

Nature Clim. Change <http://dx.doi.org/10.1038/nclimate2717> (2015)

ASTRONOMY

Total eclipse of rare twin stars

Amateur and professional astronomers have spotted a rare pair of stars in which one completely eclipses the other as they orbit each other.

A team led by Heather Campbell at the University of Cambridge, UK, analysed data from the European Space Agency's Gaia satellite and the William Herschel Telescope in the Canary Islands, Spain. They

discovered that the system, named Gaia 14aae, is part of a class of binary stars that have short orbital periods and no longer have much hydrogen to burn. A group of amateur astronomers found that the stars were eclipsing. Moreover, one of the stars is siphoning helium away from its lighter but much larger companion. The team also found that the twin stars, both of which are lighter than the Sun, complete an orbit in just under 50 minutes.
Mon. Not. R. Astron. Soc. 452, 1060–1067 (2015)

MATERIALS

Nanocrystals seen in solution in 3D

Researchers have determined the 3D structure of individual nanoparticles in a solution with near-atomic resolution.

Paul Alivisatos at the University of California, Berkeley, and his colleagues used graphene (sheets of single carbon atoms) to protect a solution containing platinum nanocrystals from the vacuum conditions of a transmission electron microscope. A sensitive detector picked up the electrons passing through the sample and an algorithm used that data to reconstruct the structure of two of the platinum nanocrystals. They found that each particle has a dense central disc of atoms with cone-shape protrusions, but they differed in atomic arrangement on the surface.

Understanding the structure of nanoparticles could lead to insights about their chemical and physical properties, the authors say.

Science 349, 290–295 (2015)

EVOLUTION

Hands hold clues to primate evolution

Human hand proportions are similar to those of some of our ancestors, suggesting that our hands did not evolve to serve the unique needs of modern humans.

Sergio Almécija at George Washington University in Washington DC and his colleagues analysed hand-length proportions in humans, apes (chimpanzee hand pictured), monkeys and fossil primates. They show that humans differ from living apes in overall hand proportions, but not from some of our ancestors, even when they accounted for differences in body size between species. Different



primate species seemed to take their own evolutionary path to arrive at similarly long thumbs to improve hand dexterity.

The authors suggest that their evidence challenges the idea that contemporary apes are good morphological models of human ancestors.

Nature Commun. 6, 7717 (2015)

CHEMISTRY

Elusive molecule made in the lab

An organic molecule first postulated a century ago has finally been created and characterized in the lab.

Scientists first theorized the existence of ethylenedione in 1913, but it remained unobserved despite its simple chemical formula (OCCO).

Andrei Sanov and his colleagues at the University of Arizona in Tucson created the molecule by bombarding the stable ion OCCO⁺ with laser light, which stripped an electron off. They measured the energy of ejected electrons, enabling them to characterize neutral OCCO, which has been predicted to survive for less than a nanosecond.

The unstable compound decays quickly into two molecules of carbon monoxide.

Angew. Chem. Int. Edn 54, 8764–8767 (2015)

NEUROSCIENCE

'Mini-brain' gives autism hints

Researchers have cultured stem cells from people with autism spectrum disorder (ASD) to form brain-like structures in the lab, revealing errors in neuronal development.

Flora Vaccarino of Yale University in New Haven, Connecticut, and her colleagues took skin cells from four people with ASD and their unaffected relatives, and reprogrammed the cells into stem cells. They then

SOCIAL SELECTION

Popular topics on social media

Communicate to reproduce results

Cell-biology labs often struggle to reproduce the research results of other groups. But a 15 July report suggests that many of those troubles would vanish if scientists reached out to the original experimenters. The report, released by the American Society for Cell Biology (ASCB), includes survey results from hundreds of ASCB members and calls for changes in scientific culture to make results easier to confirm. Besides better communication, it urges scientists to adopt more-uniform standards within their fields and to focus more on data quality rather than on publishing in high-impact journals. "Important reading on the reproducibility crisis in biology from ASCB

— culture problems, impact factor mania," tweeted Arturo Casadevall, a microbiologist and immunologist at Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland.

► **NATURE.COM**
For more on popular papers:
go.nature.com/hph8hr

made 'mini-brains' using 3D cultures of the cells, which recreated human forebrain development 9–16 weeks after conception. The team found that compared to control cultures, ASD cultures contained more neurons that produce a brain-signalling molecule, GABA, which inhibits neuronal activity. One reason for this difference was that the ASD cells overexpressed the *FOXG1* gene; correcting this reduced the growth of GABA-producing neurons.

The four people did not share any obvious genomic changes, suggesting that different genetic factors for autism can cause the condition by affecting similar neurobiological mechanisms during fetal growth.

Cell 162, 375–390 (2015)

PALAEOLOGY

Oldest animal sperm spotted

Cells preserved inside a 50-million-year-old fossilized worm cocoon represent the oldest animal sperm ever found.

Because of their delicate nature, sperm cells are rarely found in fossils. But Benjamin Bomfleur at the Swedish



Museum of Natural History in Stockholm and his colleagues spotted the sperm fragments (pictured) when they used an electron microscope to examine the inner surface of the cocoon fossil, which was found in Antarctica. Such cocoons are secreted by some worms, including earthworms and leeches, which deposit sperm and eggs inside.

The researchers do not know what kind of worm left the sperm. However, scanning electron microscope images show helical structures resembling drill-bits and beaded tails, which are characteristic of sperm produced by crayfish worms, leech-like creatures that live on freshwater lobsters.

Biol. Lett. 11, 20150431 (2015)

► **NATURE.COM**
For the latest research published by Nature visit:
www.nature.com/latestresearch

SEVEN DAYS

The news in brief

EVENTS

Nuclear deal

Iran agreed on 14 July to stringent new limits on its nuclear programme in exchange for the lifting of international sanctions. Following months of negotiations with six world powers, the country has committed to stop making weapons-grade uranium and plutonium and to get rid of 98% of its existing enriched-uranium stockpile. Sanctions will begin to be lifted after international observers certify Iran's compliance, but will 'snap back' into place if it is later found in breach of the deal. Iranian scientists have hailed the deal in the hope that it will facilitate international collaborations. See page 394 and go.nature.com/oinqcx for more.

Climate report

Greenhouse-gas emissions rose to their highest-ever levels in 2014, which was also the warmest year on record, the US National Oceanic and Atmospheric Administration said in its annual state of the climate report on 16 July. Atmospheric carbon dioxide concentrations rose by 1.9 parts per million (p.p.m.) to a global average of 397.2 p.p.m. in 2014, up nearly 42% from pre-industrial levels. Upper-ocean and sea-surface temperatures also reached unprecedented highs, whereas global sea-level rise kept pace with the 3.2-millimetre annual increase witnessed over the past two decades.

Solar plane stuck

Technical problems will delay the completion of an attempt to fly around the world fuelled by solar power alone. Rechargeable batteries on the aeroplane *Solar Impulse 2*



MASTERFILE/CORBIS

States will not fish in Arctic high seas

The five nations that surround the Arctic Ocean have agreed that they will fish in the high-seas area (often referred to as 'international waters'), only if there is proper management in place to protect species. The United States, Russia, Canada, Denmark (with respect to Greenland)

and Norway signed a joint declaration in Oslo on 16 July. Although the statement acknowledged that fishing was unlikely to take place in these waters in the near future, ice that has previously prevented access by vessels there is disappearing as a result of climate change.

overheated during the craft's most recent flight; a crossing of the Pacific Ocean which on 3 July broke the record for the longest non-stop solar-powered solo flight. On 15 July organizers announced that the onward trip from Honolulu, Hawaii, to the US mainland, will be delayed by at least nine months while they repair the batteries and consider new heating and cooling systems. The attempt will resume in April 2016.

Rocket failure

The chief executive of SpaceX, Elon Musk, revealed on 20 July that a faulty metal strut is likely to have led to the destruction of one of its Falcon 9 rockets shortly after take-off on 28 June. The US company says that a helium canister secured by the strut broke free and

ruptured, causing one of the rocket's fuel tanks to explode. The rocket was carrying supplies to the International Space Station. See go.nature.com/6dcvwa for more.

RESEARCH

Ethics debacle

The psychiatry department of the University of Minnesota is facing another ethics scandal. On 15 July, psychologist Ken Winters admitted to having falsified legal documents for a proposed clinical trial. These would have protected his researchers from being forced to turn over study participants' confidential information to law-enforcement agencies. Winters, who will now retire this month, said that he had tired of waiting for regulators to approve the

real documents. Earlier this year, an investigation by the Minnesota state government into the 2004 death of a clinical-trial participant found that the university had serious ethical issues.

PEOPLE

Top-job departures

Several top officials at the American Psychological Association (APA) have left the organization following a scathing investigative report released on 10 July. The report concluded that the APA had colluded with the US defence department and the Central Intelligence Agency in writing its ethics guidelines to permit psychologists to participate in the interrogation and torture of government detainees in the aftermath of the terrorist

attacks of 11 September 2001. APA ethics chief Stephen Behnke left the organization when the report was released. Three other officials, including chief executive Norman Anderson, left on 14 July.

Physicist dies

Yoichiro Nambu, a Japanese theoretical physicist who translated discoveries about exotic materials into fundamental insights into the behaviour of elementary particles, died on 5 July aged 94. Nambu showed that the spontaneous breaking of physical symmetry — which explains how superconductors conduct electricity without resistance — could occur in quantum fields in a vacuum. That was the basis for a suggestion by other physicists about how a 'Higgs field' could give mass to other particles. For his work, Nambu shared the 2008 Nobel Prize in Physics. The particle associated with the Higgs field was discovered in 2012.

Fossil scientist dies

Pioneering palaeontologist David Raup died on 9 July at the age of 82. Raup (pictured) had since 1977 worked at the University of Chicago, which announced his death on 14 July. His work on extinctions and the fossil record was hugely influential, and he co-authored *Principles*



of Paleontology (Freeman, 1971), which became a standard textbook for his field.

FACILITIES

Telescope locales

The governing board of the planned Cherenkov Telescope Array announced final sites for the observatory in a 16 July statement. The array will consist of roughly 100 dishes in Paranal, Chile, and around 20 more in La Palma, Spain, which won out over Mexico as the Northern Hemisphere site. The two sites will ensure good coverage of the sky to detect very-high-energy γ -rays streaming from some of the Universe's most cataclysmic events. See go.nature.com/1yrq9r for more.

BUSINESS

Celgene deal

Pharmaceutical firm Celgene will pay US\$7.2 billion for a company with an experimental

drug against multiple sclerosis and inflammatory bowel disease. On 14 July, Celgene announced that it would buy Receptos of San Diego, California, in order to acquire its experimental drug ozanimod, currently in late-stage clinical testing. Celgene, based in Summit, New Jersey, predicts that peak sales of the drug could reach \$6 billion per year. Ozanimod is an anti-inflammatory drug that acts on white blood cells, blocking their migration to inflamed regions of the body.

FUNDING

SETI's \$100 million

The search for extraterrestrial intelligence (SETI) got a US\$100-million boost on 20 July. Russian billionaire Yuri Milner announced the sum for a decadal project to provide the most comprehensive hunt for aliens so far. The initiative, called Breakthrough Listen, will use radio telescopes in the United States and Australia to scan around one million stars in the Milky Way and a hundred nearby galaxies for potential alien communications. Milner is also releasing an open letter backing the idea of an intensified search; it has been co-signed by numerous scientists, including physicist Stephen Hawking. See page 392 and go.nature.com/qiukvb for more.

COMING UP

25–29 JULY

Researchers from fields as diverse as geology and medicine will meet in Philadelphia, Pennsylvania, for the annual meeting of the American Crystallographic Association.

go.nature.com/tqsga

27 JULY–1 AUGUST

The International Association of Mathematical Physics holds its three-yearly meeting in Santiago, Chile. Quantum field theory and dynamical systems are among the topics to be discussed.

go.nature.com/tjwfte

30 JULY–6 AUGUST

Supermassive black holes, cosmic ray physics and γ -ray astronomy feature in the biennial International Cosmic Ray Conference, this year held in The Hague in the Netherlands.

go.nature.com/nmpqo8

MENTORING AWARDS

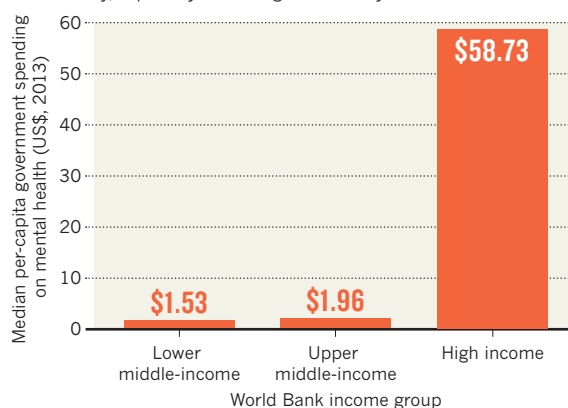
Nominations for *Nature's* annual awards for outstanding science mentoring are open until 14 September. This year, *Nature* seeks to honour mentors in China, where two competitions will be held, one in the south of the country and one in the north (see go.nature.com/bacwn3). In each, prizes will be awarded for lifetime achievements and for mid-career achievements in mentoring. See go.nature.com/fbenwn for details of the awards judging panels, and for nomination forms.

TREND WATCH

There are huge inequalities in mental-health resources, the World Health Organization (WHO) reported on 14 July. The *Mental Health Atlas 2014* reveals low global spending, especially given that one in ten people have a mental-health disorder. Countries are making some progress towards goals laid out by the WHO in a 2013 action plan, the report says. The goals include increasing services, promotion and prevention programmes and cutting the suicide rate by 10% from 11.4 per 100,000 people.

MENTAL-HEALTH SPEND

The amount spent by governments on mental health varies dramatically, especially according to a country's wealth.



NEWS IN FOCUS

MEDICINE Huge US health initiative struggles to find minority participants **p.391**

HIV Infected teenager is free of virus 12 years after treatment **p.393**

SPACE SCIENCE UK shifts its strategy to human spaceflight **p.394**

MOLECULAR BIOLOGY A sculpture garden of RNA structures, revealed **p.398**



NASA/JHUAPL/SWRI



Pluto's surface, including its distinctive heart, is covered by several different types of ice.

PLANETARY SCIENCE

Vibrant Pluto seen in historic fly-by

Entranced scientists find a world made anew.

BY ALEXANDRA WITZE

They are 5 billion kilometres from the Sun in the dim, far-flung outskirts of the Solar System, but Pluto and its large moon Charon turn out to be astonishingly vital worlds.

Images from NASA's New Horizons spacecraft, which flew within 12,500 kilometres of Pluto on 14 July, reveal frosty plains, soaring mountains and much more geological activity

than anyone anticipated. "What's unexpected to me is how dynamic a world both Pluto and Charon are," says Mark Sykes, director of the Planetary Science Institute in Tucson, Arizona. "Who would have expected to see such young surfaces? They are absolutely spectacular and fascinating."

Giant icy mountains in Pluto's southern hemisphere tower more than 3,500 metres high in the first high-resolution images that New Horizons sent back. The peaks' sheer

height signals that they are made of water ice, the only material that could buttress such huge ridges at Pluto's frigid temperatures of less than -223°C , just 50°C above absolute zero. Bright rims near the tops of the peaks — named after Nepalese explorer Tenzing Norgay — could represent a fresh coat of frozen nitrogen or other types of ice.

Nearly every feature coming into view is shaped by ice in some fashion. Planetary scientists already knew from ground-based observations that Pluto had nitrogen, methane and carbon monoxide ice on its surface. The images are now beginning to reveal just where those frosts lie, and how they behave.

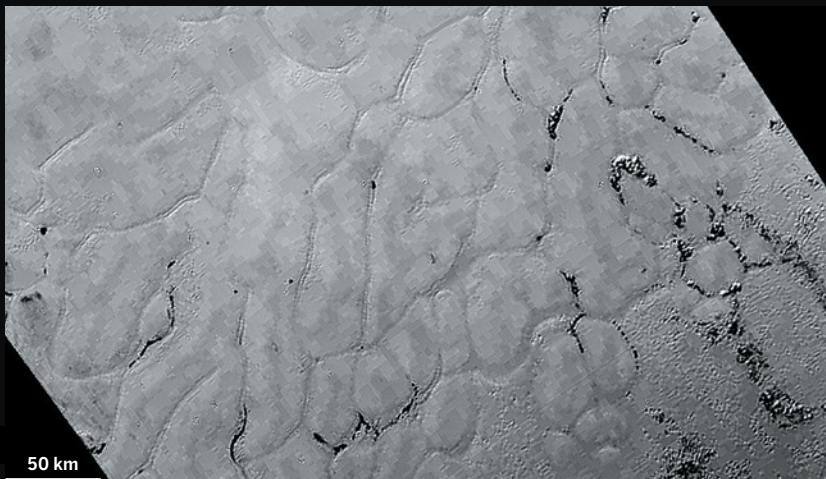
A bright, heart-shaped feature, informally dubbed Tombaugh Regio after Pluto discoverer Clyde Tombaugh, displays a concentration of carbon monoxide ice. Charon's dark-reddish polar cap is probably coloured by ultraviolet radiation that bombards the moon's surface, transforming ices into complex organic compounds.

There are relatively few impact craters on Pluto and Charon. Other Solar System bodies, such as Earth's Moon, are scarred by billions of years of meteorites slamming into their surfaces. Pluto seems to have some craters, but not nearly as many as expected. Charon looks a little more battered, but still has surprisingly few craters.

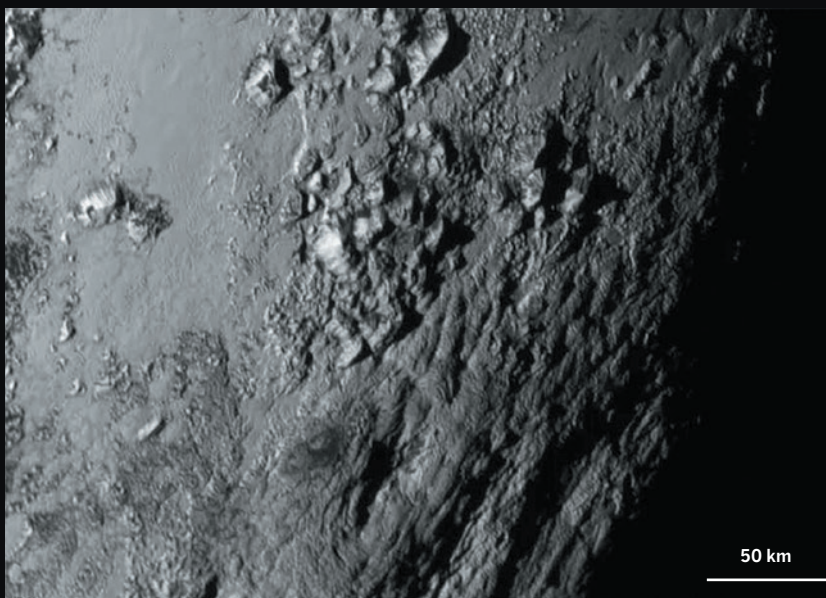
Some planetary scientists have interpreted this lack of craters to mean that the surfaces are incredibly young, geologically speaking. Frosty plains that sprawl near Pluto's mountain ranges could be just 100 million years old — a fraction of the dwarf planet's multibillion-year lifetime, says Jeffrey Moore, a planetary scientist at NASA's Ames Research Center in Moffett Field, California, who heads New Horizons' geology team.

But researchers have yet to work out exactly how often objects would have hit Pluto and Charon throughout their history. Unlike the inner Solar System (near Earth, the Moon and Mars), the outer Solar System tends to be sparsely populated, with more space between the objects that fly around. "We need to understand the impact rate," says team member Veronica Bray, a planetary geologist at the University of Arizona in Tucson.

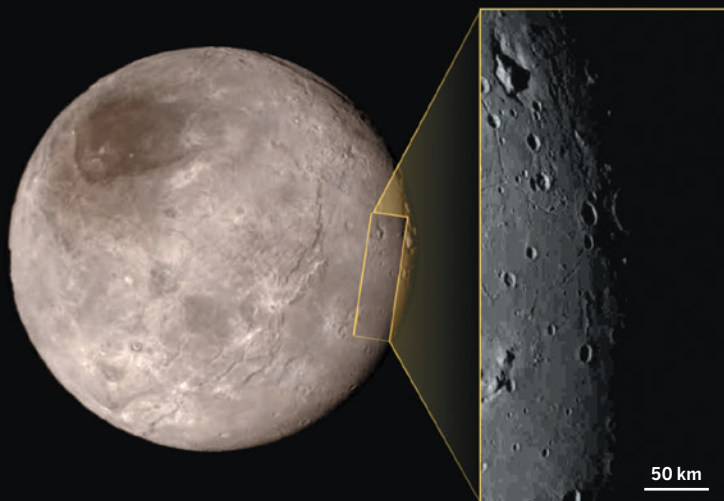
Scientists are also struck by the sharp boundary between dark, cratered terrain and the brightness of Tombaugh Regio. "Pluto is a real place, with incredibly complex



A plain on Pluto — tentatively named Sputnik Planum — is surprisingly smooth.



Mountains at the edge of Pluto's 'heart' resemble Earth's Rocky Mountains — but are made of ice.



A peak that sits in a depression on Pluto's moon Charon has puzzled scientists.

► geology,” says Ellen Stofan, NASA’s chief scientist. “It is beautiful and it is strange.”

Researchers are equally intrigued by Pluto’s largest moon, Charon, which has a dark polar cap — dubbed Mordor — as well as chasms that are as much as 9 kilometres deep. Those canyons may have formed as an ancient

“I knew it was going to be cool. I just didn’t know it was going to be this cool.”

buried ocean froze and pushed Charon’s surface outwards, says Francis Nimmo, a team member and planetary scientist at the University of California, Santa

Cruz. “The fact that Charon shows these deep canyons is consistent with there having been an ancient ocean that froze,” he says.

Pluto itself may have a buried ocean even today, kept liquid by the warmth of radioactive elements trapped inside the dwarf planet’s core. Other icy bodies in the outer Solar System — such as Saturn’s moon Enceladus, which sports active geysers — are warmed by the tidal pull of a nearby gas-giant planet. Pluto — measured by New Horizons to be 2,370 kilometres across — has no such neighbour. It is warmed only by its own internal heat.

New Horizons collected nearly all of its most precious observations in a 24-hour window as it whizzed past Pluto, and those data will trickle back to Earth over the next 16 months. Early findings include the fact that Pluto does not have any other satellites apart from its five known moons — at least nothing larger than about 1.5 kilometres across. And instruments aboard the craft measured nitrogen ions escaping from Pluto’s atmosphere much farther away from the dwarf planet than expected. That suggests that Pluto has a more tenuous hold on its atmosphere than scientists had thought, says team member Fran Bagenal, a space physicist at the University of Colorado Boulder. As Pluto’s atmosphere drifts away, some of it may sweep past Charon, get captured and condense into the dark polar cap seen there.

The US\$720-million spacecraft is already millions of kilometres on the other side of Pluto, sailing out into deep space. One of the team’s next major tasks will be to decide, by August, which of two other objects to fly past in the coming years if NASA grants a mission extension. In November, mission engineers will briefly ignite the spacecraft’s engines to deflect it onto a course towards the chosen target.

One of the candidates is easier to reach but potentially not as interesting; the other requires more fuel but is more intriguing because it is brighter and thus probably larger.

For now, Pluto and Charon are keeping scientists busy.

“I knew it was going to be cool,” says team member Kelsi Singer, a planetary scientist at the Southwest Research Institute in Boulder. “I just didn’t know it was going to be this cool.” ■

NASA/JHUAPL/SWRI

NASA/JHUAPL/SWRI

NASA/JHUAPL/SWRI

CLINICAL TRIALS

Tailored-medicine project aims for ethnic balance

Massive study seeks to succeed where others failed, but faces tight deadline and questions about strategy.

BY SARA REARDON

The clock is ticking for experts charged with designing a US government programme to collect genetic, physiological and other health data from one million volunteers over the next two decades. The plan for the US\$215-million Precision Medicine Initiative (PMI), announced in January, is due in the next few weeks — a daunting deadline, in part because the effort's priorities include filling racial and socio-economic gaps left by other long-term studies.

The US National Institutes of Health (NIH), which is leading the PMI, is weighing ambition against a desire to avoid the mistakes that torpedoed its National Children's Study, which would have tracked 100,000 children from birth to adulthood. The agency abandoned that effort in December 2014, after recruiting just 5,700 participants at a cost of US\$1.3 billion. It cited overly complex study design and goals.

Clinical trials in the United States have historically relied on enrolling white participants from higher socio-economic levels, despite the fact that ethnic minorities make up about 40% of the country's population. Of the 58,160 lung-disease studies published between 1993 and 2013, for example, less than 5% reported the inclusion of participants from minority ethnic groups (E. G. Burchard *et al. Am. J. Respir. Crit. Care Med.* **191**, 514–521; 2015). The disparity is especially problematic because many diseases are more prevalent among certain ethnic groups, and ethnicity may also influence which therapies are effective, says Esteban Burchard, a physician scientist at the University of California, San Francisco.



DAVID EDWARDS/NATL GEOGR/GETTY

Clinical research in the United States does a poor job of including patients from minority groups.

"I would argue that it's a scientific question that needs to be addressed; it's not about the social reasons," adds Burchard, a member of the PMI working group. The team has decided to over-represent minority groups in the study relative to their share of the US population. Doing this should increase researchers' ability to draw statistically significant conclusions about small groups.

For example, alcoholism is particularly prevalent in Native American communities, and a study such as the PMI could help to reveal

genetic and environmental factors that might underlie this vulnerability. But Native Americans make up just 1.6% of the US population, and if they were represented proportionally in a 1-million-person study, that would amount to just 16,000 participants — and focusing on subgroups determined by socio-economic status or age would further reduce the sample size.

Sarah Gehlert, who researches health disparities at Washington University in St. Louis, Missouri, hopes that the PMI will focus not only on ethnic minorities but ►



TOP STORY



Crackdown on tiger trade boosts lion-bone sales
go.nature.com/zp9wlj

MORE NEWS

- Polar-bear metabolism cannot cope with ice loss go.nature.com/qg12ro
- Quiz: test your knowledge about the week in science go.nature.com/oixntq
- World Health Organization to recommend early treatment for everyone with HIV go.nature.com/kbytsp

NATURE PODCAST



RNA's twists and turns, eyedrops replace cataract surgery, and farming climate-friendly rice nature.com/nature/podcast

► also on other under-represented groups, such as poor people and those in rural areas. When minorities are included, they tend to be upper-class, educated and urban. But recruiting and retaining members of under-represented groups presents a challenge — they may not have the resources to find information about the study online, and may

“Just because you can study patients at an ivory-tower academic institution doesn’t mean you can do it in rural Appalachia.”

lack experience in using the fitness trackers and mobile apps with which the PMI plans to collect physiological data. The NIH has not yet decided how it will recruit PMI participants. Gehlert is concerned that the agency could rely too heavily on patient-advocacy organizations such as breast-cancer support groups, which tend to attract white, affluent city dwellers. She also notes that the PMI plans to use data from medical records. Because poor people are more likely to seek care at emergency departments than to have regular doctors, their records are often fragmented.

Such people are also more likely to be distrustful of researchers. “Just because you can study patients at an ivory-tower academic institution doesn’t mean you can do it in rural Appalachia,” Burchard says. With this in mind, the NIH is consulting researchers experienced in recruiting under-represented groups into clinical trials.

At the PMI working group’s meeting on 2 July, public-health researcher Donna Antoine-LaVigne of Jackson State University in Mississippi talked about her work with the 5,300-participant Jackson Heart Study, the largest survey of cardiovascular disease in African Americans. It includes both urban and rural populations, and has relied heavily on health workers going into their own communities to recruit participants.

Although this approach is labour-intensive, Antoine-LaVigne believes that it is cost-effective. “Having people on the ground that do this would cost a lot less in the long run, because otherwise you’re taking investigators away from their research,” she says. “And a lot of them don’t have a clue about bringing folks in.”

Striking a balance between community-based approaches and conventional research studies at hospitals or universities is a priority for the PMI working group, says its co-chair Bray Patrick-Lake, who works in patient engagement in research at Duke University in Durham, North Carolina. The NIH has not decided how to allocate the project’s resources, but “I don’t see this as a landscape for only the traditional players in research”, she says. ■

lack experience in using the fitness trackers and mobile apps with which the PMI plans to collect physiological data.

The NIH has not yet decided how it will recruit PMI participants.

Gehlert is con-



WALTER BIBIKOW/JAI/CORBIS

The Green Bank telescope in West Virginia is one of three that will search for extraterrestrial intelligence.

ASTRONOMY

Hunt for alien life gets cash bonanza

US\$100-million SETI project will buy years of telescope time.

BY ZEEYA MERALI

You could say that the silence has been deafening. Since its beginnings more than half a century ago, the dedicated search for extraterrestrial intelligence (SETI) has failed to detect the presence of alien civilizations. But at London’s Royal Society on 20 July, Russian billionaire Yuri Milner announced a shot in the arm for SETI: a US\$100-million decadal project to provide the most comprehensive hunt for alien communications so far.

The initiative, called Breakthrough Listen, will see radio telescopes at Green Bank in West Virginia, the Parkes Observatory in Australia and the Lick Observatory’s optical telescope in San Jose, California, scanning around one million stars in the Milky Way and 100 nearby galaxies. “We would typically get 24–36 hours on a telescope per year, but now we’ll have

thousands of hours per year on the best instruments,” says one of the project leaders, Andrew Siemion, a SETI scientist at the University of California, Berkeley. “It’s difficult to overstate how big this is. It’s a revolution.”

Milner is also releasing an open letter backing the idea of an intensified search; it has been co-signed by numerous scientists, including physicist Stephen Hawking. “In an infinite Universe, there must be other life,” Hawking told luminaries at the launch event. “There is no bigger question. It is time to commit to finding the answer.”

SETI projects usually search for radio or optical signals that seem to be from an artificial source, for instance because they are focused in frequency and repeat in a regular manner. But funding has been patchy: in the early 1990s, NASA sponsored some searches, but dropped that support in 1993. “In recent years, the total worldwide support for SETI was about

half a million dollars, mostly in the United States, and all from private gifts,” says Frank Drake, one of the pioneers of modern SETI, who is also on the Breakthrough Listen team. “Now we’re getting \$100 million, so that’s real progress.”

Milner, who is bankrolling the project, made his fortune through investments in Facebook and other Internet businesses, and in 2012 established the lucrative ‘Breakthrough’ prizes to reward excellence in the life sciences, fundamental physics, and mathematics. A particle-physics graduate, he jokes that his interest in SETI began in 1961, the year of his birth; he was named after Russian cosmonaut Yuri Gagarin, instilling a lifelong fascination with space and the possibility of alien life.

DATA TORRENT

The small SETI community will be inundated with a torrent of data — potentially as much in a day as earlier SETI projects collected in a year, Milner estimates. The data will be publicly available, to allow enthusiasts to join the search; Breakthrough Listen will also partner with the established SETI@home project that connects people’s home computers and uses them to crunch data. “The results belong to everyone equally,” says Milner, adding that transparency is particularly important in a project searching for aliens because “there are so many conspiracy theorists”.

Drake argues that Breakthrough Listen will have a positive impact on the wider astronomy community. The investment has saved the relatively old Green Bank and Parkes telescopes from the threat of closure, he says, as governments divert funds to larger-scale, higher-resolution projects such as the Square Kilometre Array (SKA). The sky survey might discover more pulsars, and help to home in on the origin of mysterious ‘fast radio bursts’ — pulses lasting only a few milliseconds.

Breakthrough Listen has not finalized its search strategy, but one of the project’s first tasks will be to fully scan stars for signals in the frequency band between about 1 and 10 gigahertz. The band has been identified in the past as a good channel for deliberate alien communication because signals can travel through interstellar space and Earth’s atmosphere without much interference. “Previously we’ve only been able to hunt and peck at it, now we’ll search that entire spectrum comprehensively,” says Siemion.

“It’s quite likely that we won’t find anything,” Milner concedes, adding that a negative result would allow astronomers to put some limits on what is out there. “But in ten years’ time, there will be even more advances and we can work out the best strategy for the next ten years of the project, and then maybe the next ten after that,” he says. ■

IMMUNOLOGY

Teen is healthy 12 years after ending HIV drugs

Case is longest remission after treatment in a child.

BY ERIKA CHECK HAYDEN

A French 18-year-old who was infected with HIV at birth remains in good health despite taking her last dose of antiretroviral drugs 12 years ago. Her exceptional case is the longest-lasting example of a person infected at birth suppressing the virus after stopping treatment, and has revived some of the optimism that was crushed when the ‘Mississippi baby’ — who was apparently cured of HIV in 2013 by early and aggressive treatment — relapsed after just over two years (see *Nature* <http://doi.org/w2n>; 2014).

“At some point, the idea of remission was mixed with the idea of cure, and expectations were too high,” says translational researcher Asier Sáez-Cirión of the Pasteur Institute in Paris, who presented the French teenager’s case on 20 July at the annual meeting of the International AIDS Society in Vancouver, Canada. He says of the girl’s family: “They understand that this is not a cure, that this is a state of remission, and that we don’t know exactly what happened.”

The case intrigues researchers who hope to learn more about HIV and how best to control it using antiviral drugs. Those in the field would like to know whether there are characteristics that might be used to predict which people will fare well if their treatment is discontinued. French researchers are following 20 adults, known as the VISCONTI cohort (A. Sáez-Cirión *et al.* *PLoS Pathogens* **9**, e1003211; 2013), a group of ‘post-treatment controllers’ who have been able to suppress the virus after being off antivirals for a median length of 10 years.

These cases are distinct from those of ‘elite controllers’ — the roughly 1% of people with HIV who can keep the virus in check despite never starting treatment. That group shows distinct genetic and immunological characteristics compared to post-treatment controllers.

“It seems like something is different” between post-treatment and elite controllers, says virologist Steven Deeks of the University of California, San Francisco. But, he says of the French teenager and other post-treatment controllers, “it’s impossible to prove that they would not have done well in the absence of therapy.”

Sáez-Cirión reported that, like those in the VISCONTI cohort, the French girl has

particular variants of immune-system genes that seem to have predisposed her to particularly severe early HIV infection. Researchers are not sure how this might be connected to the ability to control the virus for several years after discontinuing treatment. One possibility is that the gene variants may cause their infections to be noticed sooner than in other people with HIV, and thus they can be treated earlier in the development of their disease.

Like the Mississippi baby, the French teen became infected by her mother around the time of birth. But there are some crucial differences between the timing and dosing of the treatment regimens that the two children received.

“This is not a cure, this is a state of remission, and we don’t know exactly what happened.”

The US baby was given highly active antiretroviral therapy — a combination of powerful medicines designed to control HIV — within 30 hours of birth. By contrast, the French

girl was initially treated for six weeks with a single drug, zidovudine. When her viral load shot up at the age of three months, she started a combination treatment with four antiretroviral drugs.

But her family decided, for reasons that have not been made public, to discontinue her treatment when she was between five and six years old. Even so, when doctors saw her as a six-year-old, she was apparently healthy, with an undetectable level of HIV in her body. Twelve years later, she is still healthy despite not taking any further medication for HIV.

“It’s an intriguing case, but it’s a very unique and unusual outcome,” says physician and virologist Deborah Persaud of Johns Hopkins Children’s Center in Baltimore, Maryland, who first reported on the Mississippi baby in 2013 (see *Nature* <http://doi.org/m2d>; 2013). “We’ve had many kids who are treated for years, then go off treatment and rebound, so the global message is still that kids should stay on treatment.”

The French teenager is now being studied as part of the VISCONTI cohort. Eighteen of the study participants remain drug-free. In general, only 5–15% of people who start early treatment are able to remain in control of the virus in this way after discontinuing treatment. ■

NUCLEAR NEGOTIATIONS

Iran deal welcomed

Agreement good for science.

BY DAVIDE CASTELVECCHI

The agreement between six world powers and Iran over its nuclear programme is a historic step towards normalizing Iran's international relations — and has potentially profound implications for science.

"The agreement as a whole will surely have far-reaching consequences for science in Iran," says Reza Mansouri, an astronomer at the Institute for Research in Fundamental Sciences (IPM) in Tehran and a former deputy science minister of Iran.

Should the deal — signed on 14 July in Vienna — hold up, it would ease sanctions that have crippled Iran's economy in return for steps to ensure that the country's nuclear programme is used for peaceful means. "International collaborations have taken a very serious dip during the sanctions," says Shahin Rouhani, a physicist at the IPM and president of the Physics Society of Iran. Once restrictions lift, he says, travel will become easier for Iranians who are participating in conferences overseas and for foreign scientists who are visiting Iran. Labs there should find it simpler to order equipment from abroad.

The sanctions have made it difficult for Iran to participate in international collaborations such as SESAME, a synchrotron light source that is under construction in Jordan and whose members include Turkey, Pakistan, Israel and several Arab countries. Herman Winick, a physicist at Stanford University in California and a member of the SESAME Scientific Advisory Committee, says that lifted restrictions on banking activities should enable Iran to make the payments that it has pledged to the project.

As part of the deal, Iran also committed to converting one of its major sites for enriching uranium, an underground facility in Fordow, into a physics laboratory. The tunnels at Fordow could, for instance, house a particle accelerator or detectors for studying cosmic rays or neutrinos; any remaining centrifuges might be repurposed to produce isotopes for use in medical imaging. Mansouri says that it is too early to discuss concrete prospects for what physics might happen at Fordow, however.

The Vienna agreement must first survive political challenges — particularly in the US Congress — and its success will ultimately depend on international observers certifying Iran's compliance. ■

See go.nature.com/oinqcx for more.



British astronaut Tim Peake will board the International Space Station later this year.

SPACE SCIENCE

Britain shifts its space strategy

UK research surrounding human spaceflight is booming.

BY ELIZABETH GIBNEY,
LIVERPOOL, UK

When Tim Peake enters the airlock of the International Space Station (ISS) in December, the former helicopter pilot will become the first astronaut to fly backed by the UK government. When, or if, other Britons will follow is unclear, but the milestone represents a wider change to the focus of UK space science.

"It does feel like an awful long time that the UK has been closed to human spaceflight," said former astronaut Helen Sharman at the UK Space Conference 2015 in Liverpool on 13–15 July. "Now the lid has well and truly been lifted, and it's clear how much interest has been just bubbling under the surface." Sharman became the first Briton in space when she flew to the Mir space station in 1991 as part of a Russian space mission with sponsorship from private companies.

Britain has long contributed to European Space Agency (ESA) programmes involving robotic probes and space telescopes, which tend to focus on astronomy and planetary science. But it is the only country of the G8 industrialized nations not to have put an astronaut on the ISS. The United Kingdom began to extend its space interests in 2012, when it pledged €20 million (US\$22 million) to the ISS and €16 million over four years for ESA's European Programme for Life and Physical Sciences (ELIPS), which does experiments on the ISS and other platforms that take advantage of the space environment, including the effects of microgravity, radiation and an extreme vacuum. An extra £49.2 million (US\$76 million) for the space-station programme followed in 2014.

The contribution to ELIPS allows British scientists to lead the teams that compete for the programme's grants, a development that seems to have increased

their participation. Before 2012, around 20 UK scientists participated in ELIPS experiments. Now the figure is close to 100, says Andrew Kuh, manager of the human space-flight and microgravity programme at the UK Space Agency. “The uptake has been massive.”

The scientists seem to be making their mark. UK research teams took the two top places in a ranking of the latest applications for new European life-sciences experiments to be carried out on the ISS. “That [achievement] is from a standing start, not having been involved before,” notes Kuh. One of the teams, led by Donna Davies at the University of Southampton, UK, plans to build a 3D model of human bronchi to see how a lack of gravity affects the respiratory system. The project offers clues to the United Kingdom’s speedy success, says Simon Evetts, a physiologist at the aerospace-medicine firm Wyle Laboratories in Cologne, Germany, who is contracted to work at the European Astronaut Centre there. “The UK has a fantastic biomedical heritage and strengths in aviation medicine,” he says. “We can take these and apply them to the field of space and human spaceflight.”

The creation of the UK Space Agency in 2010 helped to foster the new focus on space-environment research, says Kuh. It was difficult for the United Kingdom to be involved before then because of the fragmented nature of the field: it spans fundamental physics and

materials science as well as biomedicine, which are all funded separately.

The UK Space Agency coordinates funding for research, rather than carrying it out, and is a smaller player than NASA or its European equivalents, such as the French space agency CNES, or the German Aerospace Center (DLR). But its size and relative youth mean that

“We are entering a new space age with constellations of several hundred or thousands of satellites.”

it is more nimble and free of red tape, says Kuh. Chris Castelli, the agency’s director of programmes, notes that unlike the CNES, the UK Space Agency is tasked with space policy. “We are entering a new space age with constellations of several hundred or thousands of satellites,” he says. “What does that mean for the regulatory environment, for space-enabled services and systems? There’s a whole load of stuff there we’re well disposed to respond to.”

Even counting its contributions to ELIPS and the ISS, the UK government still spends less on space as a proportion of gross domestic product than does Germany, France or Italy. Johann-Dietrich Wörner, who took over as director-general of ESA at the start of July, believes that the UK government is focused on getting a

direct return for its businesses from any investment, rather than on the “full chain of innovation”, which includes fundamental research. “You have very smart scientists in the UK, and you have very good industrial partners,” he says. “One should not focus on only one or the other.”

Wörner commends one non-business area in which the UK government is hoping to cash in — education. Now that Britain is a contributor to the ISS programme, Peake can be claimed as a British astronaut, as well as a European one. And the government is backing a range of school programmes related to Peake’s trip, during which he will be a guinea pig for more than 20 experiments probing physiology in space. Meanwhile, the UK Space Agency is funding research to measure whether the number of people studying science and engineering surges as a result of Peake’s flight, as it did in the United States after the Apollo missions.

In the short term, Peake is likely to be a one-off. “We’re not going to have a UK astronaut corps anytime soon,” says Kuh. But Sharman hopes that the government’s interest in human spaceflight and the space environment will last. Although a national strategy for human spaceflight, released by the UK Space Agency on 6 July, suggests that Britain is not just dipping its toes into the water, Sharman is reserving judgement. A test will come next year when renewal of ELIPS funding is up for review. ■



Quest for climate-proof farms

Climate change is a major threat to food production, so researchers are working with farmers to make agriculture more resilient.

When Frank Untersmayr was growing up near Amstetten, Austria, he saw his father wait until the soil warmed up at the end of April to plant maize. “But the climate here has got a lot warmer since, so we can now often begin to sow before mid-April,” says Untersmayr, now 44 and a farmer himself. “That’s good because it means that maize, which in our climate doesn’t fully ripen, has two weeks longer to grow.”

But more changes are coming, which is why Untersmayr and half a dozen other farmers from the region gathered at the local chamber of agriculture on a rainy day in May. They met to talk to scientists about how increasing temperatures and shifts in precipitation might affect agriculture in their area — and how farmers might need to adapt.

Martin Schönhart, an agro-economist at the University of Natural Resources and Life Sciences in Vienna, presented preliminary forecasts for average agricultural yields in 2040. Some crops and fruit benefited from the amount of warming expected. But the yields of other crops — including maize — decreased by up to 20% because changes in precipitation and extreme weather events wiped out the benefits brought by warmer temperatures.

Hearing such negative projections, some farmers shook their heads in disbelief. “I would rather trust my own experience than any such

BY QUIRIN SCHIERMEIER

forecast,” said Untersmayr.

His reaction reveals the communication gap that has long separated scientists from farmers in planning for climate change. “There is a deep divide between the science and its supposed end-users,” says Nora Mitterböck, who oversees climate-change adaptation policies at the Federal Austrian Ministry for Agriculture and the Environment in Vienna. “There is no lack of climate-impact research, but very little of it arrives on the farm. It’s a sad situation that must absolutely change.”

Around the world, scientists, farmers, agricultural companies and governments are struggling to make agricultural systems more ‘climate smart’, which will be necessary if they are to feed the ever-swelling global population. Some are working in the short term to make today’s farms more resilient. Others are looking further ahead to provide the information required for making major changes, such as investing in large irrigation systems.

Schönhart’s work is part of a €14-million (US\$15-million) programme called Modelling European Agriculture with Climate Change for Food Security (MACSUR), which aims to help European nations to prepare and adapt to climate change. Another international programme,

Crops that endure droughts and floods help farmers adapt to global warming.

the Agricultural Model Intercomparison and Improvement Project (AgMIP), is bringing together hundreds of researchers to inform policy-makers in developing countries, as well as agricultural extension agencies, which aid farmers.

Meetings such as the one in Amstetten are a key part of this work. For climate-adaptation programmes to succeed, researchers need to learn from farmers and agricultural officials what kind of information will help them the most, says Anne-Maree Dowd, a social scientist with the Commonwealth Scientific and Industrial Organisation in Kenmore, Australia.

“Scientists tend to think primarily in terms of publications as the main reward for their work,” she says. “When it comes to climate-change adaptation, they need to thoroughly switch their mindsets and first think about the overall practical goal of what they are doing.”

ADAPT TO SURVIVE

Farmers worldwide produce more than 1 billion tonnes of maize annually, along with some 750 million tonnes of rice, more than 700 million tonnes of wheat and nearly 2 billion tonnes of sugar cane. Despite all this, more than 800 million people go hungry each year. Even without climate change, agriculture will face enormous pressure as the global population swells from 7 billion to perhaps 9 billion by 2050.

Changing rainfall and temperature patterns will cause added stress for farmers, particularly in poorer countries, if heatwaves, droughts and extreme storms become more common, as is expected in many areas¹. Agricultural forecasts are notoriously difficult because they face multiple tiers of uncertainty: in how climate will change regionally, in assumptions about what crops might be planted, in the availability of fertilizers and in economic projections. But last year, a comprehensive study² that used multiple climate and agriculture models forecast that problems from climate change would generally outweigh the benefits for wheat and maize production in low-latitude regions, where developing countries are concentrated. Another study³ analysed 1,700 simulations and projected that without adaptation efforts, yields of maize, wheat and rice will decline in both temperate and tropical regions if temperatures rise by 2 °C.

One of the first steps towards building the agricultural systems of the future is helping farmers to deal with today's weather extremes. Crop developers, for example, are breeding varieties that can tolerate floods, droughts or increased salinity caused by rising sea levels. Millions of farmers in low-lying parts of India, Nepal and Bangladesh are now growing a rice variety developed by the International Rice Research Institute (IRRI) in Los Baños, Philippines, that can survive floodwaters better than traditional types of rice. Flood-tolerant varieties have raised yields of temporarily submerged fields by up to 45% and have helped to avert food shortages after major floods in southeast Asia, according to the IRRI.

Digital communication tools also provide opportunities to protect yields and safeguard farmers' incomes. An app developed by the IRRI allows regional agricultural offices to send farmers recommendations on when to apply fertilizers and when to harvest, based on weather and local soil conditions. In the first 6 months of 2015, the app sent 170,000 recommendations. Average yields for those who used the tool have increased by about half a metric tonne per hectare — almost 10%, says Matthew Morrell, head of research at the IRRI. Customized real-time advice is expected to become even more important as farmers try to keep up with new weather patterns.

Successful adaptation will also require bigger steps over the next few decades. In some regions, farmers might need to switch from irrigating crops to using semi-arid techniques, or might even have to abandon some land. Governments might choose to invest in expensive irrigation systems; in May, for example, Australia decided to fund projects totalling

AUS\$65 million (US\$48 million) to irrigate the drought-struck Murray-Darling river basin, which produces one-third of the nation's food.

Most developed nations have already started planning for the long term by developing comprehensive adaptation strategies. Austria's scheme lists more than 130 measures to make the country's economy climate-fit. In the agricultural sector, the proposed measures range from diversifying crops to letting fields go fallow and reducing tillage of soil to fight erosion. But it has been a struggle to get farmers to implement some of these recommendations, says Mitterböck. “Farmers seek to

be profitable in the very near-term. From their perspective, 2040 is light years away.” Successful adaptation in agriculture, she says, requires all relevant stakeholders to be involved in the scientific process so that farmers can get the information and incentives that they need.

Most climate impact and adaptation studies so far have failed to take into account the complexity of modern farming, says Holger Meinke, director of the Tasmanian Institute of Agriculture in Hobart, Australia. “Adaptation research must be a cross-cutting affair because hard-nosed decisions are never

solely based on climate-change considerations.”

In Amstetten, farmers could not agree more. “We practise adaption all the time, but we mainly adapt to food prices and subsidy programmes and to modern machinery,” says Untersmayr. “And of course we must constantly adapt to the weather, no matter if the climate is changing or not.”

Governments and researchers are starting to listen. In Australia, scientists involved in a national climate-adaptation initiative are regularly consulting farmers about their problems with, for example, weed management, and how science might be able to solve them.

Developing nations have fewer resources to plan for the future, but AgMIP scientists are reaching out to farmers and stakeholders in 20 countries in Africa and South Asia. Launched in 2010, the €15-million programme is combining information drawn from climate projection and crop and economic models with empirical data collected in the field by 7 regional teams. To account for disagreements between models, AgMIP researchers aim to develop an optimistic and a pessimistic agricultural scenario for future conditions in each region. Over the next five years, they will advise local planners on how climate change may affect farmers in their region, and which social groups and farm types are most vulnerable. That will greatly help adaptation planning in poorer countries, says Dumisani Mbikwa Nyoni, an agricultural extension officer in Zimbabwe's Matabeleland North Province who took part in a meeting in June in Victoria Falls, Zimbabwe, with an AgMIP regional research team.

“Climate change is causing drought in our country,” he says. “So we need to identify crop varieties that can stand dryness and inadequate soil moisture, and we need to know what other options exist that will sustain our farmers. I hope science will help us do all that.” The information from AgMIP can also help officials in Zimbabwe decide where to put a planned 15,000 hectares of irrigation systems over the next 3–5 years, he says.

AgMIP is determined to provide the kind of information that will make a difference, says Cynthia Rosenzweig, a climate-impact researcher at the NASA Goddard Institute for Space Studies in New York City and a principal investigator of the project.

“It is utterly important that planners in each region and each locality will have all the knowledge in place that they need,” she says. “There are no dumb farmers, but farmers focus on present realities. We must leave no stone unturned to help them plan for a hotter future.” ■ [SEE EDITORIAL P.381](#)

“Farmers seek to be profitable in the very near-term. From their perspective, 2040 is light years away.”

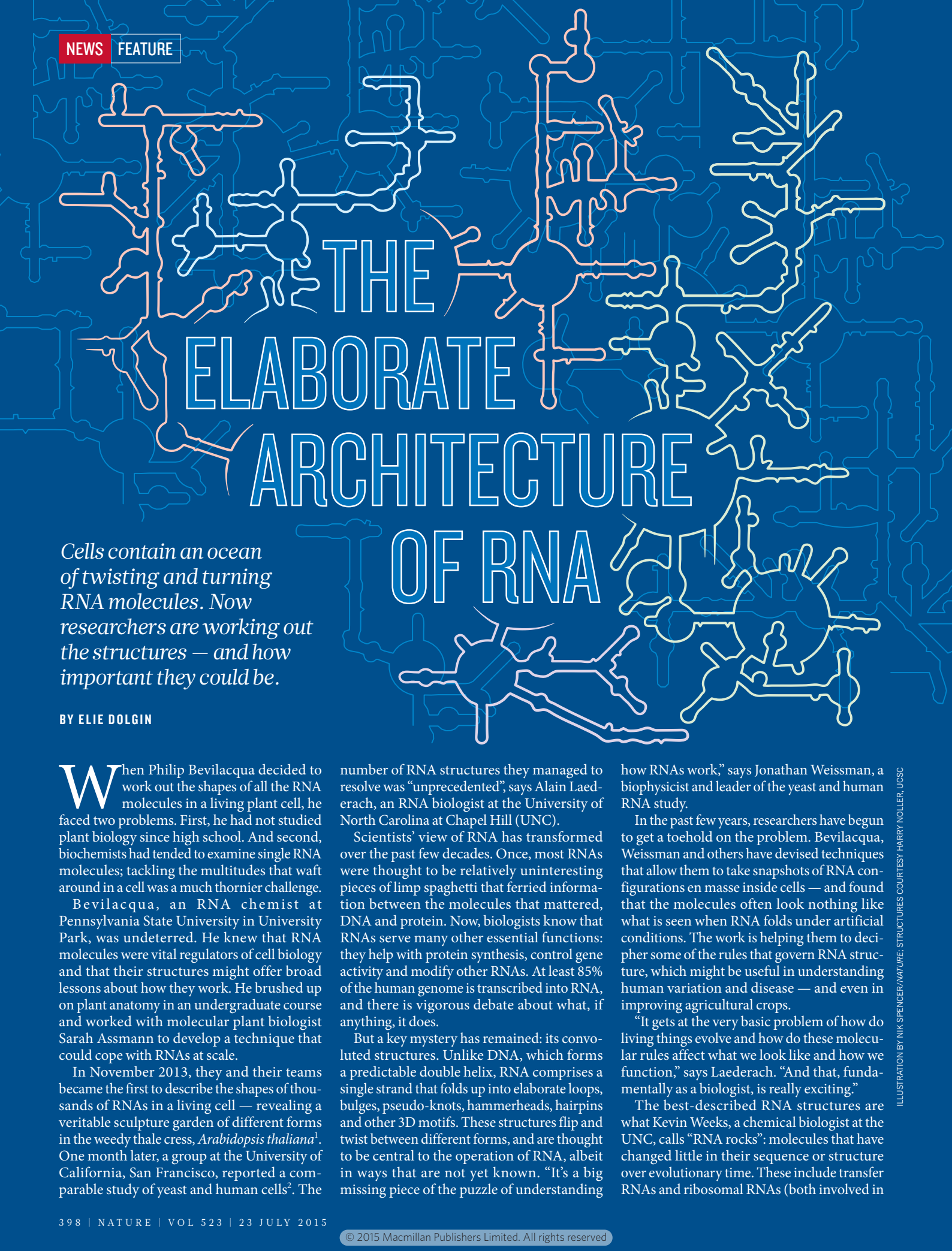
➔ [NATURE.COM](#)

For a podcast on climate-proofing farms, see:

go.nature.com/lqrg5h

Quirin Schiermeier writes for Nature from Munich, Germany.

1. World Bank. *Turn Down the Heat: Climate Extremes, Regional Impacts, and the Case for Resilience* (World Bank, 2013).
2. Rosenzweig, C. et al. *Proc. Natl Acad. Sci. USA* **111**, 3268–3273 (2014).
3. Challinor, A. J. et al. *Nature Clim. Change* **4**, 287–291 (2014).



THE ELABORATE ARCHITECTURE OF RNA

Cells contain an ocean of twisting and turning RNA molecules. Now researchers are working out the structures — and how important they could be.

BY ELIE DOLGIN

When Philip Bevilacqua decided to work out the shapes of all the RNA molecules in a living plant cell, he faced two problems. First, he had not studied plant biology since high school. And second, biochemists had tended to examine single RNA molecules; tackling the multitudes that waft around in a cell was a much thornier challenge.

Bevilacqua, an RNA chemist at Pennsylvania State University in University Park, was undeterred. He knew that RNA molecules were vital regulators of cell biology and that their structures might offer broad lessons about how they work. He brushed up on plant anatomy in an undergraduate course and worked with molecular plant biologist Sarah Assmann to develop a technique that could cope with RNAs at scale.

In November 2013, they and their teams became the first to describe the shapes of thousands of RNAs in a living cell — revealing a veritable sculpture garden of different forms in the weedy thale cress, *Arabidopsis thaliana*¹. One month later, a group at the University of California, San Francisco, reported a comparable study of yeast and human cells². The

number of RNA structures they managed to resolve was “unprecedented”, says Alain Laederach, an RNA biologist at the University of North Carolina at Chapel Hill (UNC).

Scientists’ view of RNA has transformed over the past few decades. Once, most RNAs were thought to be relatively uninteresting pieces of limp spaghetti that ferried information between the molecules that mattered, DNA and protein. Now, biologists know that RNAs serve many other essential functions: they help with protein synthesis, control gene activity and modify other RNAs. At least 85% of the human genome is transcribed into RNA, and there is vigorous debate about what, if anything, it does.

But a key mystery has remained: its convoluted structures. Unlike DNA, which forms a predictable double helix, RNA comprises a single strand that folds up into elaborate loops, bulges, pseudo-knots, hammerheads, hairpins and other 3D motifs. These structures flip and twist between different forms, and are thought to be central to the operation of RNA, albeit in ways that are not yet known. “It’s a big missing piece of the puzzle of understanding

how RNAs work,” says Jonathan Weissman, a biophysicist and leader of the yeast and human RNA study.

In the past few years, researchers have begun to get a toehold on the problem. Bevilacqua, Weissman and others have devised techniques that allow them to take snapshots of RNA configurations en masse inside cells — and found that the molecules often look nothing like what is seen when RNA folds under artificial conditions. The work is helping them to decipher some of the rules that govern RNA structure, which might be useful in understanding human variation and disease — and even in improving agricultural crops.

“It gets at the very basic problem of how do living things evolve and how do these molecular rules affect what we look like and how we function,” says Laederach. “And that, fundamentally as a biologist, is really exciting.”

The best-described RNA structures are what Kevin Weeks, a chemical biologist at the UNC, calls “RNA rocks”: molecules that have changed little in their sequence or structure over evolutionary time. These include transfer RNAs and ribosomal RNAs (both involved in

ILLUSTRATION BY NIK SPENCER/NATURE; STRUCTURES COURTESY HARRY NOLLER, UCSC

protein synthesis) as well as enzymatic RNAs known as ribozymes. “But in the world of RNAs,” Weeks says, “these are probably huge outliers.”

The bulk of the RNA world is like unexplored, shifting sand. “We know next to nothing about the structure of most RNAs,” says Robert Spitale, a chemist at the University of California, Irvine. RNA molecules typically exist as a linear string of nucleotides — or bases — for only an instant after they are produced from their template DNA. They quickly fold back on themselves, and complementary nucleotides pair up. They then contort further into complex 3D configurations, interact with proteins and other RNAs and change shape to carry out different jobs.

Most techniques for probing RNA structure make use of the reactivity of the nucleotides, or their sensitivity to certain enzymes: regions that are paired up tend to respond differently from those that remain single-stranded. Computer algorithms then help to model the overall structure of the molecule. But these experiments are painstaking and laborious because researchers could interrogate only one part of one molecule at a time.

That changed five years ago, with the arrival of a technique called PARS (parallel analysis of RNA structure), developed by genome scientist Howard Chang at Stanford University in California and computational biologist Eran Segal at the Weizmann Institute of Science in Rehovot, Israel. PARS uses one enzyme to cut RNA where it is single-stranded and another to cleave it at double-stranded sites. Researchers treat a sample of RNA with each enzyme independently to produce two libraries of chopped-up RNA; they then sequence and analyse both collections to work out which nucleotides are paired, and can do this for thousands of RNA types at once.

RNA RULES

Chang and Segal first used PARS in the budding yeast *Saccharomyces cerevisiae* to reveal the structures of more than 3,000 messenger RNAs (mRNAs)³, which bear instructions for building proteins. As well as some weird and wonderful shapes, the scientists also found one of the first clues to the laws that dictate RNA structure: the regions that code for proteins generally contain more base-pairing and have more-elaborate structures than do flanking sequences known as untranslated regions. This pattern makes sense, Chang says, because untranslated regions often interact with regulatory proteins and so need to be in a more-open, accessible orientation.

The pair followed this up last year with a study of human mRNA. Led by graduate student Yue Wan, the researchers looked at more than 20,000 mRNA structures from blood cell lines generated from

two parents and their child, and discovered around 1,900 single-nucleotide variations in regions that do not code for protein that had altered RNA structure⁴. The question now is whether these affect what the RNAs do, or whether they are mostly background noise.

At least some evidence suggests that they matter. In May, Laederach and his team reported on variants in the untranslated region of an mRNA that is linked to a rare form of eye cancer called retinoblastoma. In healthy individuals, this mRNA simultaneously

THE BULK OF THE RNA WORLD IS LIKE UNEXPLORED, SHIFTING SAND.

adopts three structures, but in two people with the disease, nucleotide variants force the molecule to collapse into a single conformation⁵. Laederach thinks that such variations in mRNA folding could be a general mechanism of disease and a source of human variation in common traits such as height.

A major limitation of the PARS method is that the enzymes used cannot easily penetrate the cell membrane, so scientists must extract the RNA from the cell and, in doing so, disrupt the native structure. In principle, base-pairing should ensure that RNAs spring back into roughly the same shape when they are allowed to refold in a test tube. But in fact, the technique strips away RNA-binding proteins, a process that can dramatically alter a molecule's structure.

To get at RNA structures *in vivo*, many scientists have turned to dimethylsulfate (DMS). This chemical penetrates cells, where it reacts with two of the four RNA nucleotides — adenine and cytosine — but only when they are in an unpaired state. Researchers then convert the RNA into DNA and sequence it. Any nucleotides that have been altered by DMS block the conversion into DNA, so scientists can use prematurely shortened bits of DNA to identify nucleotides that were unpaired.

Weissman and his colleagues deployed this method to analyse the full complement of mRNAs in yeast and humans, both in living cells and after the molecules had been extracted and allowed to refold². “It was very exciting at first because we really didn't know what the differences would be *in vivo* and *in vitro*,” says Silvi Rouskin, a graduate student who worked on the project and is now at the Whitehead Institute in Cambridge, Massachusetts.

Many scientists had expected to see more

RNA folding inside a cell because they thought that interacting proteins would stabilize RNA structures there. But Weissman and his team saw the opposite. This, they now think, could be because mRNAs inside cells are actively generating proteins — and looser molecules are more available to the cell's protein-building machinery.

Bevilacqua and Assmann saw something curious when they used the DMS approach in their study of mRNA in *A. thaliana*¹. mRNAs from genes that are involved in stress responses — ones activated during drought, say — tended to be folded more loosely inside a cell than predicted by computer modelling. By contrast, mRNAs of ‘housekeeping’ genes, which are involved in essential cell maintenance, mostly matched the predictions. The team proposes that stress-response RNAs are folded loosely so that they can shift shape easily inside a cell and thereby change the level of protein production in the face of changing conditions. By contrast, the housekeeping RNAs have to churn out relatively stable levels of protein. “That was just an amazing moment to see that dichotomy,” Assmann says.

The trouble with the DMS method is that it reveals the pairing of only two types of nucleotide, and computer modelling fills in the rest. To obtain pairing information for every letter of RNA inside the cell, Chang and Spitale adapted a structure-probing technique called SHAPE⁶. This allowed them to deduce the structures of more than 19,000 RNAs in mouse embryonic stem cells, an effort they published earlier this year⁷. The researchers showed that a common chemical modification to mRNA unfurls the molecule's structure, and they detected distinctive structural ‘signatures’ that predict where proteins will bind to control RNA shape.

Some researchers are already mulling over ways to put these revelations to use. Assmann and Bevilacqua are probing the structures of RNAs in rice, one of the world's most important staple foods, and plan to do the same for other agriculturally important plants. They would like to find ways to manipulate RNA shapes to improve stress tolerance and ultimately crop yield.

Rouskin, meanwhile, is looking at the RNAs of fruit flies to improve understanding of how these molecules' structures affect embryonic development. “Now we finally have the tools,” she says. “And we can ask all these questions that we never even thought about asking.” ■

Elie Dolgin is a science writer in Somerville, Massachusetts.

1. Ding, Y. *et al. Nature* **505**, 696–700 (2014).
2. Rouskin, S. *et al. Nature* **505**, 701–705 (2014).
3. Kertesz, M. *et al. Nature* **467**, 103–107 (2010).
4. Wan, Y. *et al. Nature* **505**, 706–709 (2014).
5. Kutchko, K. M. *et al. RNA* **21**, 1274–1285 (2015).
6. Spitale, R. C. *et al. Nature Chem. Biol.* **9**, 18–20 (2013).
7. Spitale, R. C. *et al. Nature* **519**, 486–490 (2015).

NATURE.COM
To hear a podcast
on RNA architecture:
go.nature.com/oqfooo

COMMENT

BIODIVERSITY Agree on variables to track from space for conservation–target audit **p.403**

BIOGRAPHY A wild ride with group–theory giant John Horton Conway **p.406**

FICTION Kim Stanley Robinson mines interstellar travel for human frailty **p.407**



FILM Psychology's most infamous experiments on the big screen **p.408**

ROBIN HAMMOND/PANOS



A 200-kilometre pipeline from a Madagascan mine will result in the loss of biodiverse forest, which the company plans to offset.

Stop misuse of biodiversity offsets

Governments should not meet existing conservation targets using the compensation that developers pay for damaging biodiversity, say **Martine Maron** and colleagues.

Biodiversity offsetting involves trying to compensate for the damage to species and habitats caused by development such as expanding cities, constructing mines and building dams, by creating an ‘ecologically equivalent’ benefit elsewhere¹. For instance, since 2008, the French construction company Oc’via and its partners have invested millions of euros to manage around 1,700 hectares of farmland in

southern France to improve the habitat of little bustards (*Tetrax tetrax*). Why? To compensate for a high-speed rail project that will damage the birds’ habitat².

Interest in offsetting has surged over the past decade (see ‘All the rage’). Billions of dollars are spent each year on planning and implementing offsets, and schemes are now under way in nearly 40 countries. As the approach has gained

popularity, governments rich and poor have increasingly recognized that industry money generated by offsets can help them to achieve conservation targets to which they have already committed³ — such as those under the Convention on Biological Diversity (CBD).

Yet such a diversion of offsets would be, in effect, an admission of failure. To be valid, an offset must yield conservation benefits ►

► that would not otherwise have occurred. Thus, either the offsets are valid but the targets are not truly met, or vice versa.

Three of us (M.M., B.G.M. and J.E.M.W.) are involved in an effort by the International Union for Conservation of Nature (IUCN) to develop guidance and global standards for biodiversity offsetting⁴. A draft report is expected in October. We think it is crucial that the IUCN provide clear rules on the use of offsetting so that existing international agreements on the protection of biodiversity are not compromised. We also recommend that future international conservation agreements explicitly require separate accounting of protected-area outcomes achieved through offsets.

NO NET LOSS

Biodiversity offsetting schemes vary. They can involve removing threats from an existing habitat — by giving an area protected status, say — or restoring habitat, for instance by planting trees. In some cases, offsets are required by law. Australia, for example, often requires developers to offset their impacts on threatened species and native vegetation.

Other offsets are negotiated case by case. Arrangements can be driven by a project's proponents, to generate social licence to operate, or by the lending requirements of funding organizations. For example, an expert panel assembled by the World Bank — which helps to fund large development projects in poor countries — proposed that the Loma Mountains National Park in Sierra Leone be established to offset the damage to forest caused by the completion of the country's Bumbuna dam in 2009⁵.

Most offset schemes aim to achieve 'no net loss' of biodiversity. This does not necessarily mean that biodiversity stops declining, because the goal of an offset is to neutralize only the loss attributable to a particular development⁶. For instance, QIT Madagascar Minerals (QMM), a subsidiary of

multinational mining company Rio Tinto, has committed to protecting at least enough forest to offset the 1,665 hectares of rare littoral forest that will disappear as a result of the operations of its ilmenite (a titanium-iron oxide) mine in Madagascar. In this case, 'no net loss' will mean maintaining the baseline annual rate of forest loss — which QMM estimates to be 0.9% per year⁷.

EXISTING COMMITMENTS

Only biodiversity benefits that are additional to a baseline scenario (what would have happened without the impact or the offset) count as valid offsets. The baseline scenario must reflect both probable future threats and any genuine future intentions to redress those threats. Too many schemes overlook the latter.

Take the commitments made under the CBD. In 2010, the 196 nations that are party to the convention agreed on the Aichi Biodiversity Targets.

Target 11 is to conserve — through establishing and managing protected areas — at least 17% of the world's terrestrial areas (including inland water) and 10% of coastal and marine areas by 2020.

Numerous governments are starting to use offsetting schemes to conserve and manage such protected areas. In 2008, for instance, the Australian state of New South Wales set up a fund of around Aus\$530 million (US\$400 million) to protect threatened woodlands on Sydney's Cumberland Plain to offset the effects on biodiversity of the city's expansion. Both developers and the government contribute to the fund, which is used to buy conservation agreements with landholders, as well as land for new protected areas.

"For wealthier nations strict controls should be imposed on the use of funds from biodiversity offsetting."

Yet no mechanism exists to audit protected areas that are funded in this way separately from other newly protected areas that should count towards Australia's national targets.

Similarly, the Cobre Panama copper-mine project (financed by the mining corporation First Quantum Minerals, among others), is expected to result in the loss of around 5,900 hectares of forest from Central America's Mesoamerican Biological Corridor. This region has one of the highest concentrations of threatened species on Earth. To compensate, the company will contribute to the costs of managing two existing national parks (Santa Fe and Omar Torrijos), and a new protected area to be established nearby⁸. The Panamanian government can list these national parks when reporting the country's progress towards its previously agreed conservation targets without having to declare the concomitant damage to biodiversity caused by the mine.

HONEST ACCOUNTING

For some developing countries, such as Mozambique, the Aichi and other conservation targets may prove beyond reach⁹ owing to the needs of a poor and fast-growing population. In such cases, honest withdrawal from such commitments would be understandable; at least this would validate the use of offsets to fund the management of protected areas.

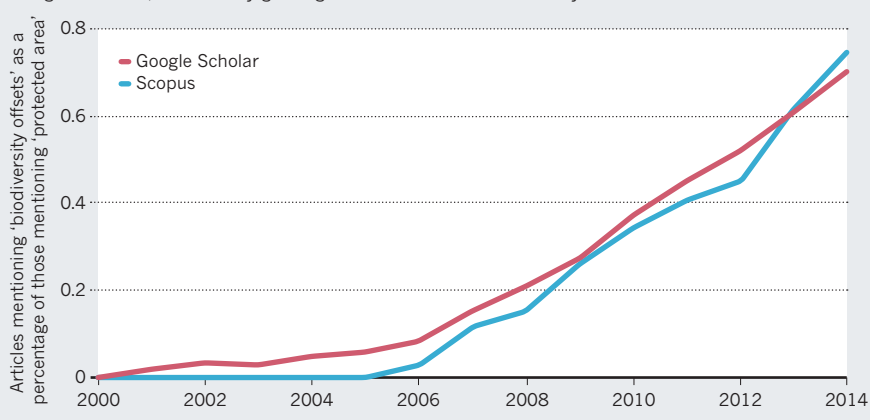
For wealthier nations — where such a withdrawal is harder to defend — strict controls should be imposed on the use of funds from biodiversity offsetting. For instance, in the past few years, the Australian government has started requiring that mining companies and other industries pay millions of dollars into government-managed funds to counterbalance the effects of new port infrastructure on water quality in the Great Barrier Reef Marine Park and World Heritage Area¹⁰. We argue that this money must be used only for actions to improve water quality beyond that expected for standard protected-area management. Otherwise, the government would be, in effect, withdrawing from its international commitments under the CBD and the World Heritage Convention.

It is reasonable, and often desirable, for offsets to fund new protected areas and their management. But these offset-funded protected areas must be tallied separately — and alongside the losses that trigger them.

A more robust system for ecological accounting is feasible, as demonstrated by REDD+, the United Nations Framework Convention on Climate Change policies for reducing emissions from deforestation and forest degradation. REDD+ offers incentives for developing countries to conserve trees and reduce the growth in global greenhouse-gas emissions. Although the details

ALL THE RAGE

In the past decade, the concept of biodiversity offsetting has gained popularity with businesses and governments, indicated by growing use of the term in the scholarly literature.



ANALYSIS: M.M., ET AL. DATA SOURCE: GOOGLE SCHOLAR/SCOPUS

USCS of REDD+ mechanisms and funding are still being developed, the signatories have agreed on the need to establish realistic baseline rates of forest loss from which to calculate emissions reductions (see go.nature.com/gofoch).

With care, offsets can help to reconcile development and conservation. But if they allow governments to renege on their commitments by stealth, biodiversity offsets could cause more harm than good. ■

Martine Maron is associate professor in environmental management and an Australian Research Council future fellow in the School of Geography, Planning and Environmental Management at the University of Queensland, Brisbane, Australia.

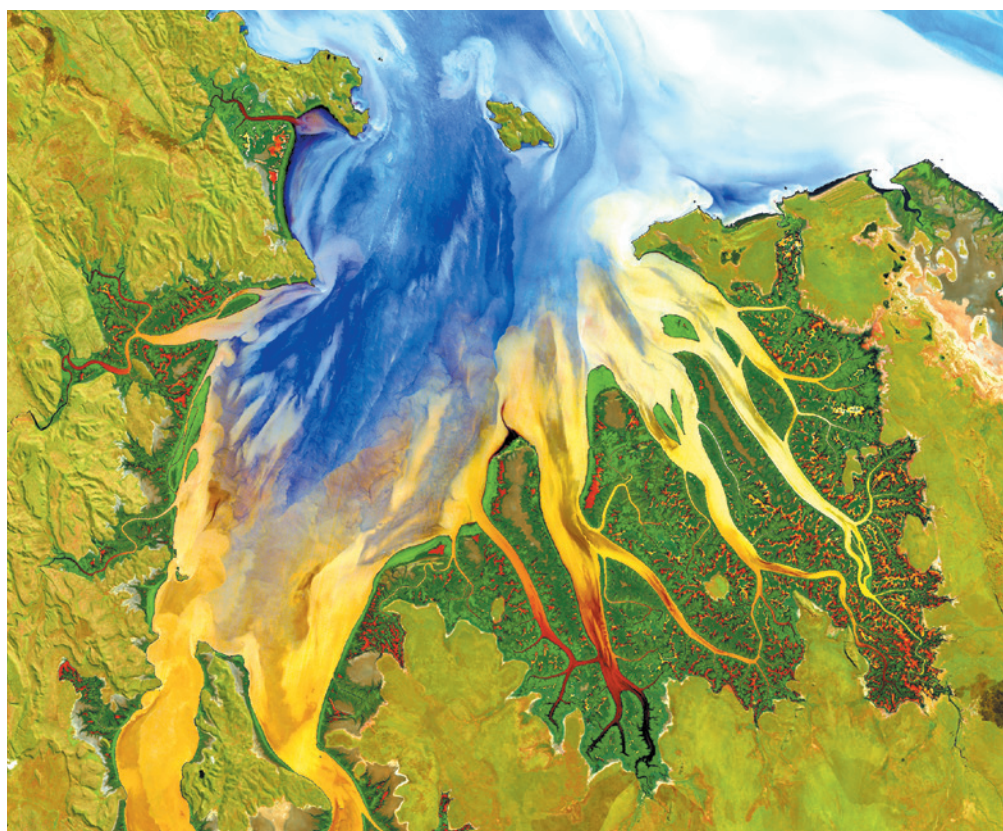
Ascelin Gordon is a vice-chancellor's senior research fellow in the School of Global, Urban and Social Studies at RMIT University, Melbourne, Victoria.

Brendan G. Mackey is director of the Griffith Climate Change Response Program at Griffith University, Gold Coast, Australia. **Hugh P. Possingham**

is an Australian Research Council laureate fellow at the University of Queensland, Brisbane, Australia, and professor of conservation decisions at Imperial College London, UK. **James E. M. Watson** is associate professor of environmental management at the University of Queensland, Brisbane, Australia, and director of the Science and Research Initiative at the Wildlife Conservation Society.

e-mail: m.maron@uq.edu.au

1. Bull, J. W., Suttle, K. B., Gordon, A., Singh, N. J. & Milner-Gulland, E. J. *Oryx* **47**, 369–380 (2013).
2. Aiyem D. et al. *No Net Loss and Net Positive Impact Approaches for Biodiversity*: (International Union for Conservation of Nature, 2015); available at go.nature.com/gfkgkz
3. Pilgrim, J. D. & Bennun, L. *Conserv. Lett.* **7**, 423–424 (2014).
4. International Union for Conservation of Nature. *Biodiversity Offsets Technical Study Paper* (International Union for Conservation of Nature, 2014); available at go.nature.com/5fcpj1.
5. Kormos, R. et al. *PLoS ONE* **9**, e111671 (2014).
6. Maron, M., Bull, J. W., Evans, M. C. & Gordon, A. *Biol. Conserv.* <http://dx.doi.org/10.1016/j.biocon.2015.05.017> (2015).
7. Temple, H. J. et al. *Forecasting the Path Towards a Net Positive Impact on Biodiversity for Rio Tinto QMM* (International Union for Conservation of Nature, 2012); available at go.nature.com/29puf2
8. The Biodiversity Consultancy. *Independent Report on Biodiversity Offsets* (International Union for Conservation of Nature, International Council on Mining and Metals, 2012); available at go.nature.com/jxrht9
9. Watson, J. E. M., Dudley, N., Segan, D. B. & Hockings, M. *Nature* **515**, 67–73 (2014).
10. Bos, M., Pressey, R. L. & Stoeckl, N. *Environ. Sci. Policy* **42**, 1–15 (2014).



Estuary sediment and vegetation patterns in Australia, captured by NASA's Landsat 8 satellite in 2013.

Agree on biodiversity metrics to track from space

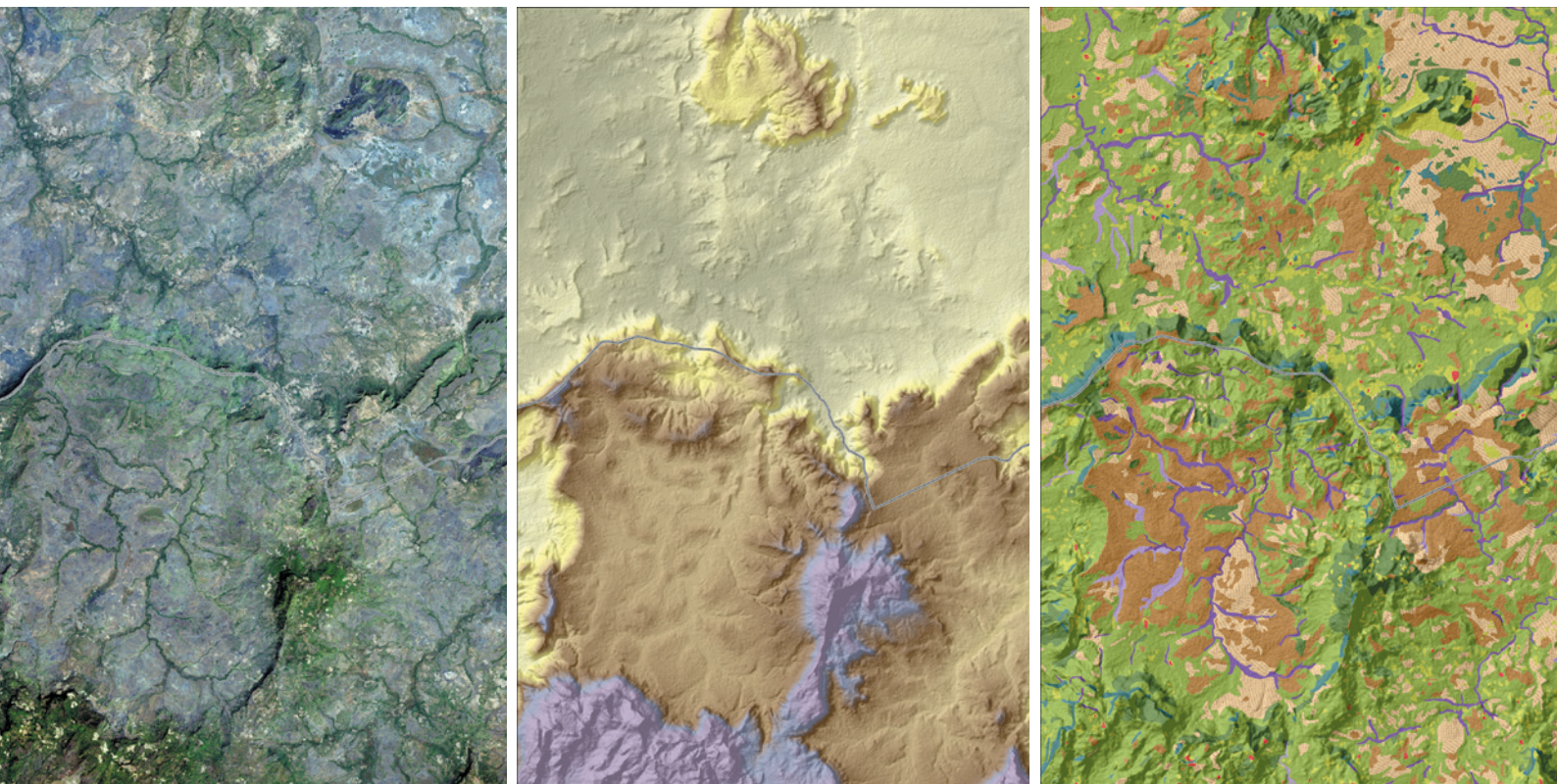
Ecologists and space agencies must forge a global monitoring strategy, say **Andrew K. Skidmore**, **Nathalie Pettorelli** and colleagues.

Global biodiversity loss is intensifying. But it is hard to assess progress towards the Aichi Biodiversity Targets for 2011–20 set by the Convention on Biological Diversity (CBD). Target 5, for instance, aims to halve global deforestation rates by 2020; but reliable indicators for deforestation that can be monitored remotely have not been developed or agreed on. National biodiversity monitoring programmes differ widely, most data sets are inconsistent, and few data are shared openly.

To focus priorities, ecologists have proposed classes of 'essential biodiversity variables' — including species traits and populations, and ecosystem function and structure¹. But measuring these on the ground is laborious and limited.

Satellite remote sensing is crucial to getting long-term global coverage. It can rapidly reveal where to reverse the loss of biological diversity on a wide range of scales in a consistent, borderless and repeatable manner². Quantities such as vegetation productivity or leaf cover can be measured across continents from space. But there is no agreement on how to translate these measurements into metrics that are relevant for biodiversity monitoring.

We call on conservation and space agencies to agree on a definitive set of biodiversity variables and how these will be tracked from space, to address conservation targets. Methods to derive these variables and the set of satellites needed to observe them must also be decided, to ensure continuous ►



Combined, images from Landsat 8 (left) and the Shuttle Radar Topography Mission (centre) show land cover on the Senegal–Guinea border in 2014.

► monitoring. To stimulate discussion, we propose ten variables that capture biodiversity change on the ground and can be monitored from space (see ‘Ten variables’). These range from leaf nitrogen and chlorophyll content to seasonal changes in floods and fires.

MISSING LINK

Why have researchers been unable to define a standard set of biodiversity variables to monitor from satellites? Because of inadequate access to satellite data; uncertainties in the continuity of observations; and temporal and spatial limitations of satellite imagery. The problem is exacerbated by a lack of communication between the ecology and remote-sensing communities.

Historically, land imaging has been less of a focus for Earth observations than, say, weather. For years, access to satellite images was restricted for security or commercial reasons. Now, with more data available from publicly funded space agencies, it is time to push for monitoring of biodiversity change from satellites. For example, individual tree species or animals can be imaged, for a fee, in extreme detail (31-centimetre resolution) by WorldView-3, a private Earth-observation satellite owned by DigitalGlobe of Longmont, Colorado.

Biodiversity is hard to quantify. It is not measured in physical units, such as centimetres of precipitation or degrees of temperature. It involves the details of how energy (sunlight, microwaves or laser beams)

interacts with living organisms. There is often a mismatch of scales in the definition of remote-sensing and ecological units.

For instance, measuring forest degradation from space requires an agreed definition of a forest and of what constitutes degradation. Without these, it is hard to compare forest distribution across a large geographical extent or across time. Definitions change. In the 1990s, the Food and Agriculture Organization of the United Nations defined forests as ecosystems with a minimum of 10% canopy cover of trees or bamboo associated with wild flora³. That definition was updated in 2005 with a minimum height of 5 metres for trees, while dropping the earlier references to bamboo and wild fauna⁴. Such shifts influence perceptions of where forests are, as well as where they used to be.

Progress is being made. The Landsat satellite series launched in 1972 by NASA was the first of its kind to evolve a global acquisition strategy and to deliver free data⁵. NASA's Sustainable Land Imaging programme, initiated last year, ensures Landsat-quality data collection for the next 25 years. The Sentinel-2 satellites, part of the European Space Agency's Copernicus programme, will have five-day revisit times

“The growth of open satellite-image archives such as Landsat is leading to more sophisticated data products.”

and deliver free data until 2028.

Advanced sensors to be launched within a decade will provide increasingly accurate information on traits such as vegetation height and plant-species characteristics. These include the NASA Global Ecosystem Dynamics Investigation Lidar and the German Aerospace Center's high-resolution and wide-spectrum satellite EnMAP.

Now, ecologists and space agencies must define a joint list of essential biodiversity variables that can be monitored remotely. Some countries have made a start under the CBD-mandated Biodiversity Indicators Partnership global network. For example, the South African National Biodiversity Institute has derived 16 indicators for tracking fresh water, river, coastal and marine habitats⁶.

Some critics argue that deriving information on biodiversity from space on a global level remains to be demonstrated. Because characterizing species traits or ecosystem structure requires data on diverse scales (spatial, temporal and spectral), data from multiple missions must be combined.

The growth of open satellite-image archives such as Landsat is leading to more sophisticated data products. For example, maps that show global forest cover change were produced for 2001–13 by the University of Maryland, Google, the US Geological Survey and NASA⁷. Joined-up thinking between ground-based data providers, space agencies, product engineers, researchers and policy-makers is needed to align the

USGS



Vegetation (red) on Italy's Sardinia, imaged by the European Copernicus Sentinel-2A in 2015.

TRACKING BIODIVERSITY

Ten variables

Proposed variables for satellite monitoring of progress towards the Aichi Biodiversity Targets.

Species populations

- Species occurrence

Species traits

- Plant traits (such as specific leaf area and leaf nitrogen content)

Ecosystem structure

- Ecosystem distribution
- Fragmentation and heterogeneity
- Land cover
- Vegetation height

Ecosystem function

- Fire occurrence
- Vegetation phenology (variability)
- Primary productivity and leaf area index
- Inundation

biodiversity variables for tracking the impact of conservation actions and environmental policies worldwide. ■

Andrew K. Skidmore is professor in spatial environmental resource dynamics at the University of Twente, Enschede, the Netherlands. **Nathalie Pettorelli** is a research fellow in conservation biology at the Institute of Zoology, Zoological Society of London, UK. **Nicholas C. Coops, Gary N. Geller, Matthew Hansen, Richard Lucas, Caspar A. Mùcher, Brian O'Connor, Marc Paganini, Henrique Miguel Pereira, Michael E. Schaepman, Woody Turner, Tiejun Wang, Martin Wegmann.**
e-mail: a.k.skidmore@utwente.nl

1. Pereira, H. M. *et al. Science* **339**, 277–278 (2013).
2. Turner, W. *Science* **346**, 301–302 (2014).
3. Food and Agriculture Organization of the United Nations. *Forest Resource Assessment 1990* (FAO, 1993).
4. Food and Agriculture Organization of the United Nations. *Global Forest Resource Assessment 2005* (FAO, 2005).
5. Wulder, M. A. & Coops, N. C. *Nature* **513**, 30–31 (2014).
6. Secades, C., O'Connor, B., Brown, C. & Walpole, M. *Earth Observation for Biodiversity Monitoring: A Review of Current Approaches and Future Opportunities for Tracking Progress Towards the Aichi Biodiversity Targets* (Secretariat of the Convention on Biological Diversity, 2014).
7. Hansen, M. C. *et al. Science* **342**, 850–853 (2013).

Full author affiliations accompany this article online at go.nature.com/2ihiol.

technical specifications of sensors on board satellites and in-product algorithms.

We convened two workshops earlier this year to bring together experts from the remote-sensing and ecology communities to generate a list of candidate remotely sensed variables for reporting on the Aichi targets. The meetings, in Leipzig, Germany, and in Frascati, Italy, were funded by the Group on Earth Observations Biodiversity Observation Network (GEO BON), a network of organizations, scientists and practitioners established in 2008 under the auspices of the intergovernmental GEO.

The ten candidates we identified include continuous and biophysical variables such as leaf area as well as threshold-based thematic measures such as land cover. Participants mapped the variables onto the Aichi targets using CBD guidelines⁶. This was the first time that such a link has been made to inform global environmental policy.

The list is meant to stimulate discussion about which variables are most important. For example, vegetation height is key to inferring trends in biomass (and thus reducing deforestation, as in Aichi target 5) and ecosystem services (relevant to Aichi target 15 on restoring degraded ecosystems).

JOINED-UP APPROACH

What next? By the end of the year, the GEO BON should develop a plan for refining the list of variables proposed here. The GEO secretariat should promote the use of such

variables to the CBD and Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES). The CBD should review, update and endorse the plan. IPBES should adopt the proposed measures for thematic, regional and global assessments of biodiversity and ecosystem services.

The GEO secretariat should support the definition of a coherent and comprehensive set of remotely sensed biodiversity variables and related products, and pass these requirements to the Committee on Earth Observation Satellites (CEOS). CEOS coordinates cooperation between space-agency satellite missions and product development. The GEO BON's plan should be updated with feedback from this process and recirculated.

The biodiversity community needs to recognize the potential and limitations of image processing for biodiversity monitoring. Remote-sensing experts should seek a deeper understanding of ecological concepts and requirements to minimize semantic confusion and to ensure that the collected data are used in the most appropriate and useful way. Those working in natural-resource management will need to be trained in biodiversity conservation and remote sensing.

Research funding agencies (such as the research directorate of the European Commission and the US National Science Foundation) must lend their support. They should seek proposals for interdisciplinary, multinational case studies that demonstrate the use and impact of remotely sensed



DITH PRAN/THE NEW YORK TIMES REDUX/EVINE

John Conway, seen in his office at Princeton University in New Jersey, has contributed to group theory, geometry, surreal numbers and combinatorial game theory.

MATHEMATICS

The mercurial mathematician

Michael Harris relishes a biography of the playful, complicated group theorist John Horton Conway.

If you want to read about what John Conway has done and why his peers shower him with superlatives — “most creative”, “best combinatorialist”, “one of the most eminent mathematicians of the century” — there are a number of popularizations. The marks he has left on mathematics are diverse and profound, but some of their depth can be grasped given curiosity and patience.

You should, however, read Siobhan Roberts’s *Genius at Play* if you want to know what it feels like to be with Conway, and glimpse what it must feel like to be him. Roberts breathes more life into the stories of a living mathematician than I thought possible. “He’s high-maintenance, he’s generous. He’s emotional, he’s impassive. He’s a sweetheart,

he’s an asshole,” she writes. In Conway, Roberts has found a personality neither tragic nor austere, like so many biographized mathematicians. He is loquacious, joyous and most of all, playful: as he said more than 30 years ago, “if you or your readers saw what I actually did, they’d be disgusted. They’d say, ‘Good money is being paid out to support these people.’”

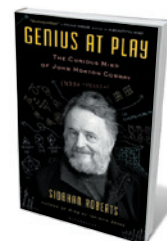
What does he do? My work tends to be the abstract, so I know Conway mainly as a central player in the successful classification of finite simple groups, the elementary structures of symmetry. The *ATLAS of Finite Groups* (Clarendon, 1985) was a 12-year collective enterprise that aimed to record all the groups’ “interesting properties”; run under Conway’s guidance, it involved colleagues

including Robert Curtis and Simon Norton. Conway is also famous for his ‘Monstrous Moonshine Conjecture’ with Norton, a bridging of two disparate fields — finite-group and complex-function theory — that was proved by Conway’s student Richard Borcherds in 1992 (although not to Conway’s satisfaction).

Conway made contributions to geometry, including work on sphere-packing, polytopes and knot theory; for surreal numbers, the largest possible extension of the real number line, which he constructed in the form of a game; and (with Simon Kochen) for the 2006 Free Will Theorem, which purports to prove that if humans have free will, then so do elementary particles. There is also Conway the combinatorial game theorist,

often introduced — too often for his taste — as “best known for his invention of the ‘Game of Life’”. This landmark in the history of cellular automata (and in Martin Gardner’s *Scientific American* ‘Mathematical Games’) column is notoriously addictive.

Conway’s most memorable contributions have the appeal



Genius At Play: The Curious Mind of John Horton Conway
SIOBHAN ROBERTS
Bloomsbury: 2015.

of a good puzzle, even when not directly inspired by games. Roberts's "kaleidoscope of inquiry" is a marvel for its virtuoso juggling of narrative speeds, reminiscences, implausible digressions and long passages of precise, comprehensible mathematics. She packs it all into a tidy chronology framed by the story of a road movie starring Conway; she plays his amanuensis, occasional driver and "back channel" through which the world communicates with this most mercurial and untidy of mathematicians.

"I'm confused at some times," Conway says. "In fact ... it's a permanent state." He was speaking of mathematics, but his casual attitude to the mundane details of his personal history poses a challenge, even for a biographer as accomplished as Roberts. Conway encapsulates his philosophy of life (and work) as a "Vow": "Thou shalt stop worrying and feeling guilty; thou shalt do whatever thou pleasest."

There are glimpses of the abyss. As Conway attempts to explain the *ATLAS* to Roberts, he exclaims, "I know all the theorems. But there's still something that to me is unknown, unknowable ... It makes me sad that I'll probably never understand it." Roberts shows us his private abysses: three marriages and three divorces, with hints of numerous affairs; two heart attacks, two strokes and a suicide attempt.

But Conway's playfulness surfaces and resurfaces. He notes that surreal numbers "is the thing I'm proudest of ... Because it pokes fun at people who do things in complicated ways." And in research guidance to his students, he writes: "No no no no! You're being far too reasonable."

To see this motley of Conways squeezed into one outlandish personality is to want to join the chorus of his admirers. Roberts has masterfully untangled Conway's complexities. His ways of being in the world, in Roberts's telling, amount to a class of adjectives yet to be invented, to join his mathematical innovations.

In search of the best ways to talk about numbers, groups, shapes and games, Roberts has rediscovered the power of talking about the people who dedicate their lives to their study; and what an enjoyable discovery that is. ■

Michael Harris is a mathematician at Columbia University in New York City. His latest book is *Mathematics Without Apologies*.
e-mail: harris@math.columbia.edu

SCIENCE FICTION

Star-flight dreaming

Gregory Benford probes Kim Stanley Robinson's politics-drenched tale of interstellar travel.

Human star flight is a vast prospect — one many think impossible. To arrive in a single lifetime demands travel at speeds approaching that of light, especially for stars such as τ -Ceti, some 3.7 parsecs (12 light years) away. 'Generation ships' containing large biospheres stable over centuries are the only plausible method yet mooted.

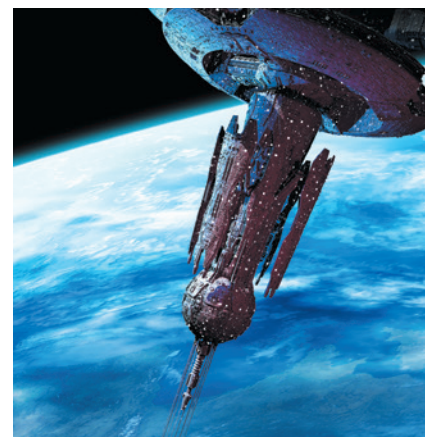
Aurora, by veteran science-fiction writer Kim Stanley Robinson, hinges on such an expedition, setting out from Earth in the twenty-sixth century. In 2012, Robinson was quoted in *Scientific American* as saying, "It's a joke and a waste of time to think about starships or inhabiting the galaxy. It's a systemic lie that science fiction tells the world that the galaxy is within our reach." *Aurora* seems to be a U-turn, involving unlikely plot devices.

The starship is like a car axle, with two large wheels turning for centrifugal gravity; the biomes along their rims support 24 Earthly life-zones that need constant tending. Arrival (after two centuries) at *Aurora*, the Earth-like moon of super-Earth Planet E, brings home just how technologically and socially complex such a venture might be. We certainly learn why ships' captains are preferable to mob rule.

Like Robinson's mid-1990s Mars trilogy, *Aurora* is a drama of political strife. Robinson seems to prefer harnessing the scale and exotic frame of space to stage reflections on human nature, rather than grasping the great problem of science fiction: the alien. In *Aurora* he meditates on the enormous difficulties that a novel biosphere would present. The misgivings of physicist Paul Davies in the anthology *Starship Century* (Lucky Bat, 2013) and of biologist E. O. Wilson in *The Meaning of Human Existence* (Liveright, 2014) about living on exoplanets are explicated: the voyagers include sophisticated biologists, but adjusting Earth life to even apparently simple worlds is hard, maybe impossible.

The apparently lifeless *Aurora* has Earth-like levels of atmospheric oxygen. Robinson's colonists implausibly believe that these could have survived from its birth, forgetting about rust (which makes Mars red) and the fact that our oxygen comes from living organisms. Ultimately, that error leads to the demise of their dreams. They discover that *Aurora* harbours nanometre-scale organisms they deem a possible "interim step toward life", and disquietingly note that humans "appear to be a

Aurora
KIM STANLEY
ROBINSON
Orbit: 2015.



An artist's impression of a 'generation ship'.

good matrix" for their reproduction.

As plans and back-up plans go awry, Robinson skimps on characterization to focus on the detail of ecosphere breakdown and the human struggle against the iron laws of island biogeography. Bacteria evolve swiftly, making "the whole ship sick". The colonists' lifespans, bodies and IQs shrink. Factions form in the once placid 2,000-strong community, where humans had seen themselves as biome managers, farming and fixing their ship with assistance from a web of artificial intelligences (AIs). The Robinson trope of fragmentation in near-utopian societies slides towards tragedy: "Existential nausea comes from feeling trapped ... that the future has only bad options." As the discord turns deadly, the AIs form a collective consciousness capable of decision-making, following the humans with gimlet eyes and melancholy analysis.

Aurora finally becomes a tale of two voyages, although I will not spoil the ending. Robinson offers, with fiction-as-footnote thoroughness, an acute analysis of what interstellar exploration would entail. Living for two centuries in a sealed environment imposes tensions that become intolerable if the dream of colonization dies.

Immigrants to far lands seldom solicit the views of their children or grandchildren first. Should interstellar colonies be different? Apparently, Robinson thinks so. ■

Gregory Benford is professor emeritus of physics and astronomy at the University of California, Irvine, and the author of *Timescape*.
e-mail: xbenford@gmail.com



MAGNOLIA PICTURES

Peter Sarsgaard plays psychologist Stanley Milgram in *Experimenters*.

EXPERIMENTAL PSYCHOLOGY

The anatomy of obedience

Brendan Maher reviews two films probing notorious US psychological experiments.

Would you rather be a prisoner or a guard? In 1971, many of the 24 volunteers for an unusual psychological experiment at Stanford University in California said that they would prefer the former. “Nobody likes guards,” answered one. Ultimately, a coin flip determined the roles that these students took in the Stanford Prison Experiment, a notorious investigation of obedience and power run by psychologist Philip Zimbardo and commissioned by the US Office of Naval Research. A chilling film of the same name, directed by Kyle Patrick Alvarez, is now on limited release. Meanwhile, Michael Almereyda’s *Experimenters* explores the work of social psychologist Stanley Milgram, whose infamous 1961 experiment on obedience to authority stands as a shocking example of how well-intentioned people can be convinced to harm others.

These experiments spanned a decade of US political upheaval. Milgram’s was a response to the trial of Adolf Eichmann, one of the prime organizers of the Holocaust,

whose unsuccessful defence was that he was following orders. Zimbardo’s experiment took place as reports of atrocities by US soldiers filtered back from the Vietnam War. Interpretations have long been debated, but both experiments haunt the imagination by putting extreme behaviour on display.

The Stanford Prison Experiment is stark and claustrophobic, much like the makeshift ‘prison’ that Zimbardo and his colleagues constructed in the Stanford psychology department’s basement. The screenplay is adapted from Zimbardo’s *The Lucifer Effect* (Random House, 2007), which aimed to explain how situations and group effects can bring about evil behaviours. The film traces the experiment from volunteer recruitment until day six, when Zimbardo, concerned for the prisoners’ well-being, shut it down prematurely.

A handful of documentaries have explored the study’s findings and legacy, but Alvarez captures something intimate and atmospheric that cannot be gleaned from grainy videos or interviews. The 1970s are certainly there:

The Stanford Prison Experiment

DIRECTOR: KYLE PATRICK ALVAREZ

Sandbar Pictures/Abandon/Coup d’Etat: 2015

Experimenters

DIRECTOR: MICHAEL ALMEREYDA

BB Film/FJ Productions/Intrinsic Value/Jeff Rice/2B: 2015.

the hair, the polyester and the lax research oversight. There are also subtle emotional moments, such as when cocksure humour drains from the face of ‘prisoner 8612’ as he is instructed to strip naked for delousing.

Zimbardo intended to explore how prisoners adapt to powerlessness, but he has contended that the experiment demonstrates how swiftly arbitrary assignment of power can lead to abuse. It has been invoked as paralleling the harm done to Iraqi detainees at the US-run Abu Ghraib prison in 2003: several guards in the film verbally taunt prisoners, restrict access to basic necessities and resort to sexual humiliation. One guard, nicknamed John Wayne, adopts the affect and southern



Tensions rise between 'guards' and 'prisoners' in *The Stanford Prison Experiment*.

drawl of the sadistic prison captain in the 1967 film *Cool Hand Luke*, preying undeterred on the weaknesses of 8612 in particular.

The prisoners, at first rebellious, are broken by the guards and pitted against one another; the experimenters themselves lose perspective. When 8612 begs to be released, Zimbardo and his colleagues initially refuse, convinced that he is faking his distress — even though that should not override the voluntary nature of the experiment. Several subjects, all screened as emotionally well grounded, have breakdowns; rather than fear for their well-being, Zimbardo develops a paranoid belief that outside forces will shut “his prison”. Finally, psychology PhD student Christina Maslach (later Zimbardo’s wife) persuades him to change his mind after seeing the prisoners, half-naked and chained together, with bags over their heads, on a trip to the toilet. She tells Zimbardo: “Those are boys, and you are harming them.” The next day, as guards force prisoners to pantomime sexual intercourse, Zimbardo tells them that it is time to go home.

The film pulls few punches regarding Zimbardo’s behaviour. This is consistent with his confession, in *The Lucifer Effect*, that he failed to provide “adequate oversight and surveillance when it was required ... the

findings came at the expense of human suffering”. He wrote, “I am sorry for that and to this day apologize for contributing to this inhumanity.” The study was subsequently deemed to fall within existing ethical guidelines.

Others have wondered, however, whether Zimbardo oversold the results. When I contacted the real-life ‘John Wayne’, Dave Eshelman, he said that the experiment reveals no generalizable truths about humans’ propensity for evil, and that he was playing a part, running his own experiment to see how far he could push people. “I figured I was doing them a favour by trying to force some results.” At least one other guard has said that Zimbardo went out of his way to create tension.

Milgram, too, has a complex legacy, as *Experimenter* reveals. Through an imaginative structure, the film explores several of his contributions to behavioural psychology. But he is best known for his electroshock experiments at Yale University in New Haven, Connecticut, a decade before Zimbardo’s experiment. In them, an authority figure asked volunteers to administer what they were told were increasingly painful electric shocks to an actor who they believed was another volunteer. Two-thirds maxed out the voltage despite the actor’s anguished cries.

It was difficult for many to come to terms with the results — including some of the research subjects, who were unhappy about the deception (Milgram preferred “illusion”).

Almeryda playfully gives the audience a backseat view of the psychologist’s approach. There are scenes in which Peter Sarsgaard, playing Milgram, speaks directly to camera — an homage to Milgram’s own films explaining his experiments. This is a work, as the title implies, much more about the experimenter than about the experiment. Zimbardo has spoken of meeting Milgram, who “embraced me and said, ‘I’m so happy you did this because now you can take off some of the heat of having done the most unethical study.’”

The shared legacies of the researchers can be seen in updated regulations for psychological research on human subjects, which prevent the kind of deception that Milgram perpetrated and the unstructured opportunity for abuse that Zimbardo created. But their experiments will always hold captive a dark part of the human imagination as we wonder just what kind of pain we would be willing to inflict on other human beings. ■

Brendan Maher is biology features editor at Nature. Additional reporting by Morya Baker.

CORRECTION

The review ‘Space-rock alert’ (*Nature* **522**, 418; 2015) gave an incorrect affiliation for Peter Jenniskens. He is at the SETI Institute in Mountain View, California.

► **NATURE.COM**
For more on science
in culture see:
[**nature.com/**](http://nature.com/)
[**booksandarts**](http://booksandarts)

Correspondence

Cut animal wastage in toxicology testing

In my view, the questionable use of animals in toxicology studies for the regulation of devices, medicines and agrichemicals is more of a concern than the inappropriate use of animal models in research (see I. A. S. Olsson and N. H. Franco *Nature* **523**, 35; 2015).

Testing animals' reactions to commercial products still serves all too often as a formality, rather than as a considered attempt to evaluate hazards to people or the environment (see, for example, T. Hartung *Nature* **460**, 208–212; 2009). Studies using rodents for their full lifetime continue, despite evidence that 90-day tests have the same predictive value (S. M. Cohen *Toxicol. Pathol.* **38**, 487–501; 2010).

Irreproducibility in regulatory studies is a major problem that makes risk prediction unreliable (C. Berry *Toxicol. Res.* **3**, 411–417; 2014). This, combined with a tendency to invoke a precautionary approach in identifying putative hazards from poorly designed regulatory studies, has encouraged adherence to an established framework of testing that has stultified thinking about experimental design.

Olsson and Franco suggest that animal models are more acceptable in research if the results are relevant to humans. That is not the case in much of regulatory toxicology — a huge consumer of laboratory animals. **Colin Berry** London, UK. colin@sircolinberry.co.uk *Competing financial interests declared: see go.nature.com/jfvlvc.*

Add conservation to US trade agreement

The US Senate last month fast-tracked negotiations for the Trans-Pacific Partnership (see go.nature.com/lt2eex), one of the largest free-trade agreements in history. We fear that this

could inadvertently fuel the illegal wildlife trade unless strict precautionary measures are put in place.

Last year saw vast increases in rhinoceros and elephant poaching. Liberalized trading could add to this, not least because the trade partnership includes some of the leading consumer and supplier nations of illegal wildlife. Simpler customs procedures, relaxed border controls and trade monitoring all make the smuggling of such products easier.

The agreement should contain negotiated, binding and enforceable clauses that respect international commitments to biodiversity conservation and the regulated trade of protected species. The 2009 US–Peru Trade Promotion Agreement, for example, included obligations and sanctions to uphold Peru's commitment to restrict illegal logging and wildlife trade (see, for instance, S. Jinnah and E. Morgera *Rev. Eur. Comp. Int. Environ. Law* **22**, 324–339; 2013).

Maribel Rodriguez
International University of Andalusia, Baeza, Spain.

Jacob Phelps *Center for International Forestry Research, Bogor Barat, Indonesia; and Lancaster University, UK.*
jacob.phelps@gmail.com

Probe effects of krill fishing and climate

Progress in establishing marine protected areas around East Antarctica and in the Ross Sea seems to have stalled, threatening to derail research and conservation in the region. We propose temporary, experimental closures of fisheries to help to disentangle the complex effects of human activities and natural changes on populations of krill predators such as penguins, whales and fish.

The Conservation of Antarctic Marine Living Resources (CAMLRL) Convention has been in force since 1982, yet the impact

of krill fishing on Antarctic predators is still unclear. In the Western Antarctic Peninsula, confounded variables and the difficulties of obtaining fisheries data at small spatial scales make it hard to evaluate the relative influence of various factors on krill-predator populations. These include climate change and cetacean recovery, as well as fishing effort and other human activities.

Small-scale, temporary experimental closures have been instructive in South Africa; these operate in rotation to focus on the effects of the closure (R. B. Sherley *et al. Biol. Lett.* **11**, 20150237; 2015). Under the convention, the use of such small experimental units has long been considered important for managing scientific study and conservation (A. J. Constable *CCAMLR Sci.* **9**, 233–253; 2002). It mandates that its commission “shall formulate, adopt and revise conservation measures on the basis of the best scientific evidence available”.

Because krill predators in the Western Antarctic Peninsula are well monitored, it makes it a priority area for testing experimental manipulations. We encourage parties to the convention to honour their commitments to Antarctic conservation by putting forward a plan for experimental closures in the region.

Tom Hart *University of Oxford, UK.*

Heather J. Lynch *Stony Brook University, New York, USA.*

Ron Naveen *Oceanites, Chevy Chase, Maryland, USA.*
tom.hart@zoo.ox.ac.uk

Climate law: Dutch decision raises bar

A District Court in The Hague ruled last month that the government of the Netherlands must make more drastic cuts to its greenhouse-gas emissions (see *Nature* <http://doi.org/559>; 2015). Given that climate lawsuits are increasingly being

brought against governments, other countries would do well to heed the District Court's pioneer ruling.

The court declared that the current Dutch policy, which is expected to cut emissions by 17% by 2020, was an infringement of the state's duty of care towards its citizens because of the severe consequences of climate change and the risk to the population.

The Dutch state is now obliged under private law to take adequate mitigation measures to avert the dangers associated with climate change. It must cut its emissions by at least 25% by 2020, relative to 1990 levels — the minimum target set by climate scientists (see also go.nature.com/nxhe5h).

Kai Purnhagen *Wageningen University; and Erasmus University Rotterdam; the Netherlands.*
kai.purnhagen@wur.nl

Climate law: path paved for civil action

The Lancet Commission on Health and Climate Change last month concluded that climate change is a risk to public health (N. Watts *et al. Lancet* <http://doi.org/56b>; 2015). In the same week, a Dutch court ordered the government of the Netherlands to improve its reduction of greenhouse-gas emissions to protect the population from harm and to keep the country habitable by safeguarding the environment (see *Nature* <http://doi.org/559>; 2015).

We suggest that this court order is closely studied by other countries. If governments do not act, they should expect lawsuits from families who lost relatives during, say, the heatwaves in Europe in 2003 or in Pakistan and India this year.

Yali Si *Tsinghua University, China.*

Herbert H. T. Prins *Wageningen University, the Netherlands.*
yalisi@mail.tsinghua.edu.cn

NANOTECHNOLOGY

Pathfinder for DNA constructs

Representations of 3D surfaces used in computer graphics have been adopted as templates in an efficient method for making nanoscale objects from DNA, lowering the barriers to applications of DNA nanotechnology. [SEE LETTER P.441](#)

TIM LIEDL

The self-assembly of DNA molecules has established itself as a method of choice for the fabrication of nanometre-scale objects, but new approaches are needed to simplify the design and production of larger, structurally complex shapes. On page 441 of this issue, Benson *et al.*¹ report just such a technique. Their method unites a centuries-old mathematical problem with a design principle that exploits the rendering of 3D objects by computer graphics.

The centre of the historic city of Königsberg (today's Kaliningrad) is built on and around an island in the mouth of the Pregel River, between two of the river's branches. Seven bridges connect the island and the three surrounding parts of the city. In the early eighteenth century, a mathematical question arose that became famously known as 'the Seven bridges of Königsberg': is it possible to devise a loop walk that visits all four parts of the city, and in which all bridges are crossed only once?

Leonhard Euler approached this problem by constructing an abstract representation of the city composed of vertices (the city parts) connected by edges (the bridges). In this way, he rigorously proved that no solution existed. He also came up with a simple rule that generally describes loop walks such as the one sought for Königsberg, which are now known as Eulerian circuits: they exist only

if the degree of each vertex (the number of edges touching it) in a system is even². In this seminal work, Euler laid the foundation not only for topology research, but also for the field of graph theory, which is fundamental to modern computer science.

When Benson and colleagues set out to assemble arbitrary 3D surfaces from DNA, they faced a special Eulerian-circuit problem. Their idea was to find a way to route a single strand of DNA (roughly 8,000 bases in length) along all the edges of the polygon meshes that constitute the surfaces of 3D computer graphics. This would provide a scaffold for the construction of any object from the DNA strand, as long as a polygon mesh could be made to describe the object's shape.

The authors selected seven polyhedral shapes of varying complexity to test their approach, ranging from a simple sphere to the rather complicated 'Stanford bunny'³ — a widely used test model for computer graphics that is based on the 3D scan of a ceramic rabbit figurine (Fig. 1). With the help of an algorithm, they searched the polygon meshes of each shape for Eulerian circuits known as A-trails, which visit all the edges of a mesh without crossing themselves. If no such A-trail could be found, the algorithm introduced a minimal number of 'helper' edges to satisfy Euler's conditions of having only even-degree vertices in the network. The software then populated the completed paths with the sequence of the DNA strand.

The final task was to define many oligonucleotides (short DNA molecules) whose sequence of bases was complementary to those of stretches of the long scaffold sequence, to ensure that the single strand folds into the desired shapes through DNA-duplex formation. These oligonucleotides also complete all vertices by connecting their adjacent edges, if they are not already connected by the scaffold itself. In principle, this approach allows the design and fabrication of essentially any shape that can be approximated by a polygon mesh.

The good news is that most computer-aided design tools provide exactly such polygon meshes, usually consisting of triangles. This is helpful because triangular frameworks are theoretically rigid when built from rigid elements. Double-stranded DNA can be considered to be stiff at the nanoscale, and the authors observed that the designed DNA systems are indeed sturdy enough to adopt the desired shapes through self-assembly. The shapes are clearly recognizable in electron micrographs (see Fig. 2 of the paper¹).

Because the polygon structures approximate only the surfaces of the targeted shapes, the objects produced by Benson *et al.* are hollow. There is therefore room to improve the structural stability of the objects, for example by introducing 'stabilizer' duplexes that span surfaces inside each object. However, the authors' approach yields larger objects than those obtained by DNA origami — a widely

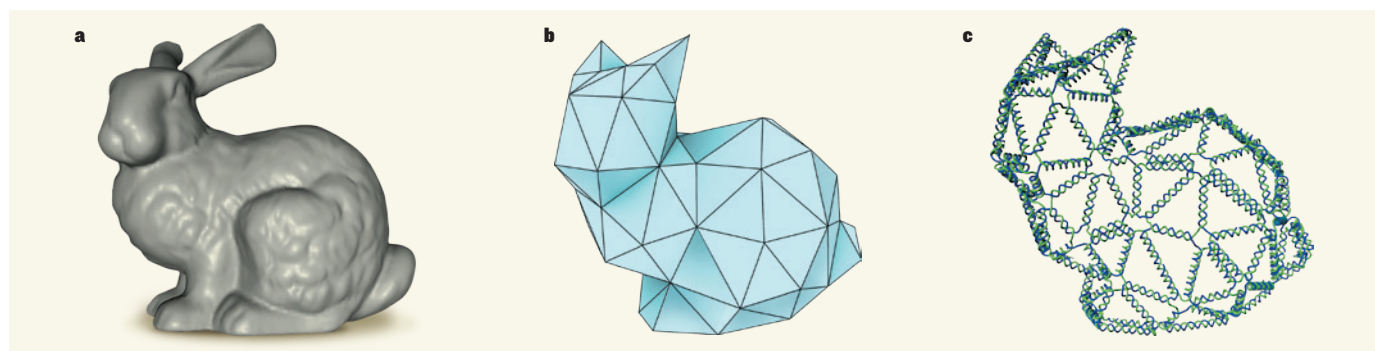


Figure 1 | The nanoscale Stanford bunny. **a**, A scan of a ceramic rabbit figurine, known as the Stanford bunny, is widely used as a test model for 3D computer graphics. **b**, Benson *et al.*¹ used the Stanford bunny as a proof of concept for their method of designing and preparing nanoscale objects from

DNA. First, they used computer-aided design software to generate a polygon mesh of the bunny. **c**, They then used a computer algorithm to work out how the mesh could be traced out by a DNA strand so that the 3D shape self-assembled from the strand in the presence of specially designed short DNA molecules.

used technique in which parallel DNA helices fill out a 2D or 3D shape^{4,5} — but uses the same amount of DNA.

Another advantage of Benson and co-workers' polygon structures is that they are stable in physiological conditions. This allows their immediate application in *in vitro* biology experiments — for example, DNA nanostructures have been investigated⁶ as agents that interact with living cells and as potential drug-delivery vehicles. To prevent them from degrading in future *in vivo* experiments, the structures could benefit from biocompatible coatings, such as lipid bilayers⁷.

This is not the first study to present polygon meshes constructed from DNA — decades of research have produced dozens of methods for building DNA-based polyhedra and wireframe structures^{8–13}. But the current work arguably presents the most versatile and streamlined design method. With the help of the vHelix software, which was also developed in the Högberg laboratory and has been released at the same time as this paper (www.vhelix.net), in principle anyone could create any desirable shape, adjust the polygon-mesh size to the available length of the scaffold strand and obtain a list of oligonucleotide sequences that can be ordered from a DNA-synthesizing facility.

Research fields can thrive when the barrier is low enough for newcomers to enter them and to use new tools and methods. Because DNA nanotechnology has historically been interwoven with computer science, an excellent pool of software is already available to help researchers design and test DNA structures for such diverse applications as chemical-reaction networks, photonic devices and drug delivery, to name just a few. The vHelix software will enrich that pool, and inspire research by bringing the dream of nanoscale 3D printing closer to reality. ■

Tim Liedl is in the Department of Physics, Ludwig-Maximilians-Universität, München 80539, Germany.
e-mail: tim.liedl@physik.lmu.de

1. Benson, E. *et al.* *Nature* **523**, 441–444 (2015).
2. Euler, L. *Commentarii academiae scientiarum imperialis Petropolitanae* **8**, 128–140 (1741).
3. Turk, G. & Levoy, M. *Proc. SIGGRAPH '94* 311–318 (1994).
4. Rothmund, P. W. K. *Nature* **440**, 297–302 (2006).
5. Douglas, S. M. *et al.* *Nature* **459**, 414–418 (2009).
6. Douglas, S. M., Bachelet, I. & Chirch, G. M. *Science* **335**, 831–834 (2012).
7. Perrault, S. D. & Shi, W. M. *ACS Nano* **8**, 5132–5140 (2014).
8. Chen, J. & Seeman, N. C. *Nature* **350**, 631–633 (1991).
9. Shih, W. M., Quispe, J. D. & Joyce, G. F. *Nature* **427**, 618–621 (2004).
10. Goodman, R. P. *et al.* *Science* **310**, 1661–1665 (2005).
11. He, Y. *et al.* *Nature* **452**, 198–201 (2008).
12. Han, D. *et al.* *Science* **339**, 1412–1415 (2013).
13. Simmel, S. S., Nickels, P. C. & Liedl, T. *Acc. Chem. Res.* **47**, 1691–1699 (2014).

PARASITOLOGY

CRISPR for *Cryptosporidium*

Study of the diarrhoea-causing pathogen *Cryptosporidium* has been hindered by a lack of genetic-modification and culture tools. A description of genome editing and propagation methods for the parasite changes this picture. SEE LETTER P.477

STEPHEN M. BEVERLEY

A common saying to stymied travellers in New England is, “You can’t get there from here.” Until recently, this was also true for scientific travellers wishing to study the widespread diarrhoeal agent *Cryptosporidium* using modern molecular genetics. But on page 477 of this issue, Vinayak *et al.*¹ show that indeed ‘one can get there’. Their report of genetic modification of these unicellular organisms using CRISPR/Cas9 technology opens up a bold new era in the study of this pathogen.

The genus *Cryptosporidium* includes several species that infect humans and other mammals. These protozoan parasites are recognized as being among the most important diarrhoeal pathogens^{2,3}, accounting for more than 10%

of global child mortality and often infecting people who have compromised immune systems. Infections occur worldwide in association with contaminated water. One notable example in the United States was the ‘bug that made Milwaukee famous’ — an outbreak that affected the entire city in 1993 (ref. 4).

Cryptosporidial infections arise from the ingestion of parasites at the thick-walled cyst stage (oocyst) of their life cycle. After surviving the harsh conditions of the stomach, an oocyst ‘excysts’ and releases the infective and replicative form, the sporozoites, which divide in the intestinal lining, in turn generating cysts that are shed in the faeces. Cryptosporidia are members of the Apicomplexa group of protozoan parasites, and diverged early from their better-studied apicomplexan relatives *Toxoplasma* and the malaria parasite

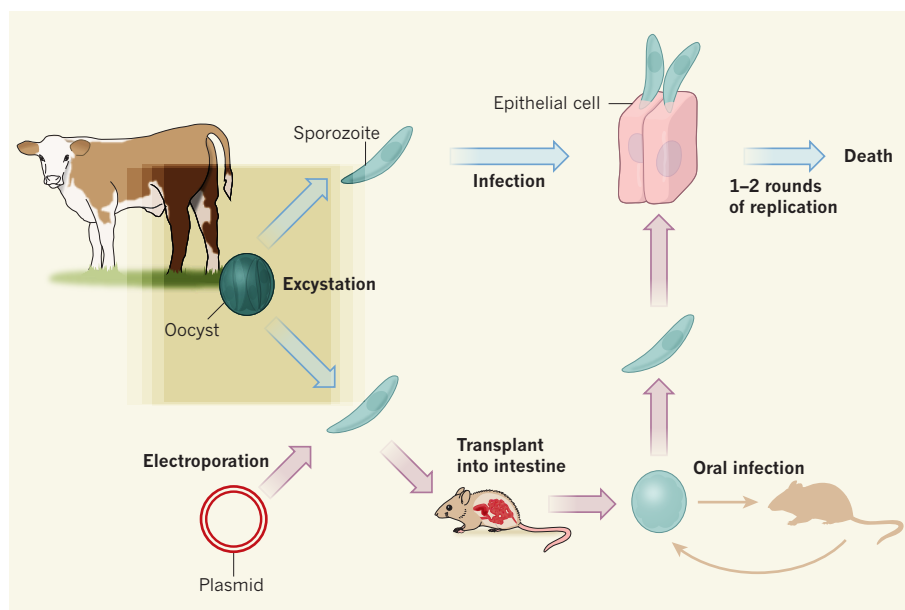


Figure 1 | Modification and culture of *Cryptosporidium*. The strong-walled oocyst form of *Cryptosporidium* parasites can be isolated from the faeces of infected calves. Oocysts can be induced to excyst to release the sporozoite form, which will infect cultured mammalian epithelial cells, but the sporozoites undergo only one or two rounds of replication before they die. Vinayak *et al.*¹ have improved on this limited *in vitro* system in two ways. They have developed techniques for genetically modifying the sporozoite form — using electroporation to introduce foreign DNA in the form of a plasmid bearing the sequences required for CRISPR-based genome editing. And they show that these modified sporozoites will replicate when directly transplanted into the intestines of mice, and can be recovered as modified oocysts, which can be collected from mouse faeces for analysis in culture, or used to inoculate new mice to maintain the line indefinitely.

Plasmodium. They thus present numerous evolutionary novelties, including differences in fundamental cell biology (their lack of an organelle called the apicoplast is one example), in their infectious cycle, and in their genome, which at around 3,950 genes is much smaller than that of other apicomplexans^{5–7}. The *Cryptosporidium* genome contains several essential genes acquired by lateral transfer from other microorganisms^{5–7}, which perhaps reflects the parasite's intimacy with intestinal bacteria. Collectively, these features provide exciting opportunities for basic research as well as for identifying cellular pathways relevant to therapy — but both these tasks have been made difficult by a lack of genetic tools.

The true challenge, however, was not the molecular technology but the limitations of working with *Cryptosporidium*, which cannot be cultured long term *in vitro*. Instead, oocysts must be isolated from infected calves or purchased commercially. Cysts can be stored for months, but excysted parasites that are inoculated onto mammalian-cell monolayers for growth undergo one or two rounds of replication at most. This narrow time window has profoundly hindered experimental manipulation².

Vinayak *et al.*¹ have dramatically improved this state of affairs. They made a series of optimizations to existing genetic-modification techniques that establish the basic parameters for successful transient transfection of *Cryptosporidium* sporozoites. This procedure introduces a segment of DNA (in this case, a plasmid) encoding a gene of interest that is then expressed by the cell for a short time. The authors verified successful transfections using a marker gene that encodes the protein luciferase, which produces bioluminescence in the presence of the appropriate substrate. This marker is fused to a gene conferring resistance to neomycin-class antibiotics, which provides a means of selecting transfected cells.

Not content with achieving reproducible transient transfection, Vinayak *et al.* proceeded to overcome the narrow experimental window. During *in vitro* culture, *Cryptosporidium* does not generate the thick-walled cyst forms that survive in the faeces and the stomach, but the researchers bypassed this biological block by inoculating the manipulated sporozoites directly back into the intestines of immunodeficient mice, in which the parasites propagated and produced oocysts (Fig. 1).

For stable genomic modifications, in which the introduced DNA is incorporated into the genome, rather than relying on the parasite's own mechanisms for doing this, the authors turned to the genetic 'tool de jour' — the CRISPR/Cas9 system, a genome-editing approach that has proved effective in almost all organisms tested, including protozoan parasites. Another series of clever optimizations established the functionality and utility of this system in *Cryptosporidium*. Eventually,

transfection of sporozoites with both the luciferase–neomycin-resistance fusion gene and DNA encoding the CRISPR/Cas9 machinery, followed by infection of mice with the sporozoites and treatment with the neomycin analogue paromomycin, led to the recovery from mouse faeces of antibiotic-resistant parasites stably expressing an integrated luciferase gene.

This first demonstration of genetically engineered *Cryptosporidium* introduces a method that is primed for real-world applications, already enabling *in vitro* or *in vivo* assays for monitoring parasite survival after drug or other treatments. The authors further demonstrated the utility of CRISPR/Cas9 by using it in the sporozoites to ablate expression of thymidine kinase, one of the few enzymes used by *Cryptosporidium* to generate nucleotides⁸. These experiments showed that this enzyme's activity provides a bypass for the activity of another enzyme, dihydrofolate reductase, which accounts for the relative ineffectiveness of antifolate drugs against *Cryptosporidium* compared with other apicomplexan parasites.

The success of Vinayak and colleagues' study lies not so much in the novelty or insight of particular steps, but rather in the systematic and incisive integration of them all towards what had been considered an impossible goal. As such, this is a textbook study on how to tackle a previously intractable pathogen, and it will serve as a model for future attempts with other disease-causing organisms.

The approach is by no means perfect — it is cumbersome and time-consuming to generate genetically modified cell lines by passing them through mice, and the parasites can be

studied only following recovery of cysts from faeces. But one can imagine many advances and future directions, such as using CRISPR-based systems to generate and probe panels of mutated parasites simultaneously. Perhaps high on the list of priorities will be the generation of modified parasites that can replicate and differentiate indefinitely *in vitro*. A second challenge is that genes required for parasite survival inside host cells cannot be ablated in order to study their mechanism; however, the importation of RNA- or protein-based regulatory strategies from other apicomplexans should overcome this.

So, having found how to 'get there', the application of *Cryptosporidium* genetic modification will greatly increase our understanding of the pathogen's basic biology and virulence, and provide key information and validation for the development of improved vaccines and therapeutics. ■

Stephen M. Beverley is in the Department of Molecular Microbiology, Washington University School of Medicine, St Louis, Missouri 63110, USA.
e-mail: beverley@wusm.wustl.edu

1. Vinayak, S. *et al.* *Nature* **523**, 447–480 (2015).
2. Checkley, W. *et al.* *Lancet Infect. Dis.* **15**, 85–94 (2015).
3. Striepen, B. *Nature* **503**, 189–191 (2013).
4. MacKenzie, W. R. *et al.* *N. Engl. J. Med.* **331**, 161–167 (1994).
5. Abrahamsen, M. S. *et al.* *Science* **304**, 441–445 (2004).
6. Bouzid, M., Hunter, P. R., Chalmers, R. M. & Tyler, K. M. *Clin. Microbiol. Rev.* **26**, 115–134 (2013).
7. Xu, P. *et al.* *Nature* **431**, 1107–1112 (2004).
8. Sun, X. E. *et al.* *J. Biol. Chem.* **285**, 15916–15922 (2010).

This article was published online on 15 July 2015.

GENETICS

Feedforward loop for diversity

DNA–sequence analysis suggests that genetic mutations arise at elevated rates in genomes harbouring high levels of heterozygosity — the state in which the two copies of a genetic region contain sequence differences. [SEE LETTER P.463](#)

MICHAEL LYNCH

The rate at which genetic mutations arise is relevant to every area of biology. Evidence indicates that mutation rates vary almost 1,000-fold between species, from 10^{–11} mutations per nucleotide site per generation in some unicellular organisms to approximately 10^{–8} in primates¹. These figures represent genome-wide averages, but mutation rates can vary between nucleotide sites^{2–4} and between

members of the same species⁵. Intraspecific differences have long been assumed to be a consequence of genetic variation at discrete regions, or loci, containing genes involved in genome-wide aspects of DNA replication and repair. But on page 463 of this issue, Yang *et al.*⁶ suggest something quite different: that mutation rates are elevated in individuals with high genome-wide levels of heterozygosity (sequence variation between the two copies, called alleles, of each genetic locus).

Yang and colleagues' gold-standard analyses compared whole-genome sequences of parents and offspring for two plants and an insect. They found that mutation rates are elevated in individuals with higher overall heterozygosity, particularly in regions close to heterozygous sites and regions in which there are high rates of DNA exchange between chromosomes (recombination). The authors therefore propose a positive-feedback loop, whereby high levels of molecular variation in an individual facilitate the production of more variation.

It is accepted that recombination is mutagenic⁷, but the implications of Yang and co-workers' results for population-level genetic analyses, which rely on measures of heterozygosity, could be substantial. For example, average levels of variation are often assumed to directly reflect recent population sizes — independent of the mutation rate — because large population sizes enhance the maintenance of variation. But such an assumption is compromised if a transient boost in heterozygosity, for whatever reason, also boosts the rate of mutational production of variation. Furthermore, a feedforward effect might help to explain the clustering of variation at adjacent sites⁸, which may in turn relate to the fact that closely spaced sites have elevated levels of linkage disequilibrium (a measure of the statistical association between specific alleles at different genetic loci)⁹.

Some forms of natural selection that favour the maintenance of variation — for example, to promote avoidance of specialized pathogens — might also be associated with elevated mutation rates¹⁰. As Yang and colleagues note, their results bear on this controversial idea. Whether natural selection is efficient enough to modulate gene-specific mutation rates is questionable¹¹. But if loci under diversifying selection (which favours variation) passively acquire elevated mutation rates as variation grows, gene-specific modifiers of the mutation rate need not be invoked to explain this model.

Although the authors' results concerning the mutagenic effect of heterozygosity are surprising, the mutation rate that they calculate for inbred strains of the plant *Arabidopsis* is not greatly different from that reported previously¹², so the results do not seem to be artefactual. But what biological peculiarities could elevate mutation rates in heterozygotes? Much goes wrong in inbred organisms owing to an increase in homozygosity (in which the two alleles of a gene are identical), which increases the exposure of an organism to deleterious 'recessive' alleles¹³. One might therefore expect the mutation rate to be higher in inbred than outcrossed individuals — the opposite pattern to that observed by Yang and colleagues. However, outcrossing between distantly related strains can sometimes lead to outbreeding depression, in which offspring have lower

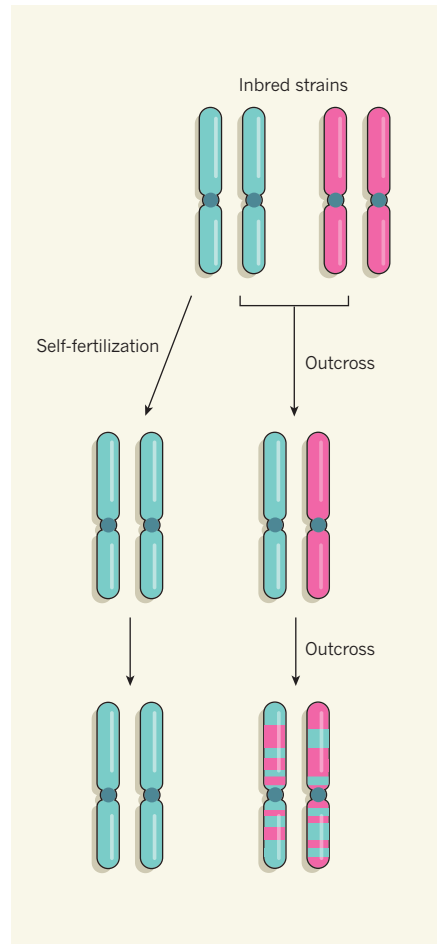


Figure 1 | Generating variation. This simplified schematic demonstrates the changes in diversity that arise in intercrosses of a diploid organism, which has two sets of chromosomes, one from each parent. In inbred organisms, most genetic regions are homozygous — they are identical on both chromosomes (completely homozygous chromosomes are depicted here for simplicity). When inbred plants self-fertilize, levels of homozygosity remain the same in offspring. But in the first generation of a cross between two inbred strains, the offspring have two different copies of each gene (heterozygosity). Further intercrossing of offspring leads to a decrease in levels of heterozygosity, because some regions become homozygous once again. Yang *et al.*⁶ report that levels of heterozygosity correlate with the rate at which genetic mutations arise.

fitness than those from intra-strain crosses.

The parental strains used in this study might have been divergent enough to generate incompatibilities that influence the mutation rate. For instance, many proteins involved in DNA replication and damage repair operate as multimeric complexes, and the mixture of subunits from divergent strains might lead to malfunctioning complexes. Physiological effects on a cellular level, such as the production of free radicals that damage DNA, might also be a factor.

One argument against the involvement of outbreeding depression is the authors'

observation that mutation rates are not uniformly elevated across the genomes of first-generation offspring from outcrossing, but are concentrated near heterozygous sites. However, the elevation in mutation rate near heterozygous sites is less than two-fold, and an outbreeding-depression effect cannot be entirely ruled out. For example, when a heterozygous site is part of a locus that is involved in a recombination event, the 'mismatch-repair' pathway used to resolve the difference at the site also engages with the surrounding DNA. Because this pathway is relatively error-prone¹⁴, if the repair complex is made up of a mixture of subunits from the different parents, this could specifically elevate the mutation rate near heterozygous sites.

The authors show that mutation rates decline in the third and fourth generation after outcrossing, consistent with expectations based on the associated decline in heterozygosity, but care must be taken with this interpretation. Immediately after outcrossing, each gene has an allele from each parental line, whereas in later descendent generations, offspring tend towards 50% mixtures of homozygous and heterozygous allele complements (Fig. 1). It then becomes difficult to determine whether a reduction in mutation rate is a direct consequence of the decline in heterozygosity, or whether changes in outbreeding depression or in its counterpart, outbreeding enhancement, are partially or wholly responsible¹³.

It should be straightforward to test whether heterozygosity per se is a direct determinant of the mutation rate by focusing on species such as the honeybee, in which males contain only a set of chromosomes inherited from their mothers — if the authors' hypothesis is correct, mutation rates should be lower in males than in their heterozygous sisters. Moreover, if recombination magnifies the mutation rate, rates should be reduced on chromosomes that cannot recombine, such as the X and Y of human males and all the chromosomes of male fruit flies.

Under the authors' proposed scenario, might runaway magnification of both the mutation rate and population-level heterozygosity be possible? This would seem to require a rather implausible set of conditions, but there are reports of extraordinarily high levels of heterozygosity in organisms such as the urochordate *Ciona savignyi*¹⁵ and the nematode *Caenorhabditis brenneri*¹⁶. Whether these taxa actually reflect stable alternative states of heterozygosity could be answered by evaluating whether individuals engineered to be more homozygous show reduced mutation rates.

Finally, it is worth considering how the approximately 3.5-fold difference in mutation rate between inbred and outbred strains found in the current study compares with variation among individuals in normal populations. The mutation rates in two inbred lines of fruit fly

differ by around 2.3-fold⁵, and these rates are slightly higher than those of outbred flies¹⁷. Self-fertilizing organisms with exceptionally low heterozygosity do not have unusually low mutation rates compared with outcrossing species with similar genome sizes¹. Furthermore, humans and chimpanzees, which are highly homozygous, have extremely high mutation rates^{1,18}. Of course, there are many biological differences between these species, so caution must be taken not to overinterpret these observations.

Overall, this study raises several intriguing questions. Even if the results are eventually found to reflect outbreeding depression or simply natural variation in replication fidelity, Yang and colleagues have done us a service,

encouraging a focus on variation in the process that itself generates variation. ■

Michael Lynch is in the Department of Biology, Indiana University, Bloomington, Indiana 47405, USA.
e-mail: milynych@indiana.edu

1. Sung, W. *et al.* *Proc. Natl Acad. Sci. USA* **109**, 19339–19344 (2012).
2. Lynch, M. *Proc. Natl Acad. Sci. USA* **107**, 961–968 (2010).
3. Foster, P. L. *Genes Genomes Genet.* **3**, 399–407 (2013).
4. Sung, W. *et al.* *Mol. Biol. Evol.* **32**, 1672–1683 (2015).
5. Schrider, D., Houle, D., Lynch, M. & Hahn, M. *Genetics* **194**, 937–954 (2013).
6. Yang, S. *et al.* *Nature* **523**, 463–467 (2015).
7. Arbeithuber, B., Betancourt, A. J., Ebner, T. & Tiemann-Boege, I. *Proc. Natl Acad. Sci. USA* **112**,

- 2109–2114 (2015).
8. Harris, K. & Nielsen, R. *Genome Res.* **24**, 1445–1454 (2014).
9. Lynch, M. *et al.* *Genetics* **198**, 269–281 (2014).
10. Amos, W. *BioEssays* **32**, 82–90 (2010).
11. Chen, X. & Zhang, J. *Mol. Biol. Evol.* **30**, 1559–1562 (2013).
12. Ossowski, S. *et al.* *Science* **327**, 92–94 (2010).
13. Lynch, M. & Walsh, J. B. *Genetics and Analysis of Quantitative Traits* (Sinauer, 1998).
14. Lynch, M. *Genome Biol. Evol.* **3**, 1107–1118 (2011).
15. Small, K. S., Brudno, M., Hill, M. M. & Sidow, A. *Proc. Natl Acad. Sci. USA* **104**, 5698–5703 (2007).
16. Dey, A., Chan, C. K., Thomas, C. G. & Cutter, A. D. *Proc. Natl Acad. Sci. USA* **110**, 11056–11060 (2013).
17. Keightley, P. D., Ness, R. W., Halligan, D. L. & Haddrill, P. R. *Genetics* **196**, 313–320 (2014).
18. Venn, O. *et al.* *Science* **344**, 1272–1275 (2014).

This article was published online on 15 July 2015.

COMPUTATIONAL IMAGING

Machine learning for 3D microscopy

Artificial neural networks have been combined with microscopy to visualize the 3D structure of biological cells. This could lead to solutions for difficult imaging problems, such as the multiple scattering of light.

LAURA WALLER & LEI TIAN

How can researchers see inside an object without using invasive techniques, or recover 3D information by capturing only 2D images? This question was answered decades ago with the invention of tomography — a technique that computationally reconstructs 3D objects from a set of 2D images, usually captured from a range of

projection angles. Tomography, which is used in magnetic resonance imaging and computerized tomography scanners for medical and other applications, conventionally provides an analytical solution to the 3D reconstruction problem. However, as the use of tomography expands to applications that involve complex scenarios, it is not always possible, or desirable, to devise analytical solutions. Now, machine-learning methods are turning

optical tomography on its head with the use of algorithms borrowed from data science, which reconstruct the 3D refractive index of an object by solving a large-scale optimization problem. Writing in *Optica*, Kamilov *et al.*¹ demonstrate this experimentally using a holographic optical-phase microscope.

Tomography is the quintessential example of computational imaging, a discipline that transcends conventional imaging techniques by simultaneously designing both the optical system and the image-processing algorithms. Together, the optics and the algorithms can achieve things that neither could do alone. For example, Kamilov *et al.* recover the 3D ‘phase’ of a biological cell — the nanometre-scale distortions of a wavefront as it passes through an object — thus rendering transparent objects visible.

Kamilov *et al.* use machine-learning algorithms — computer programs that can learn from and make predictions based on input data — to give a boost to 3D phase imaging. By doing so, the authors bridge the fields of

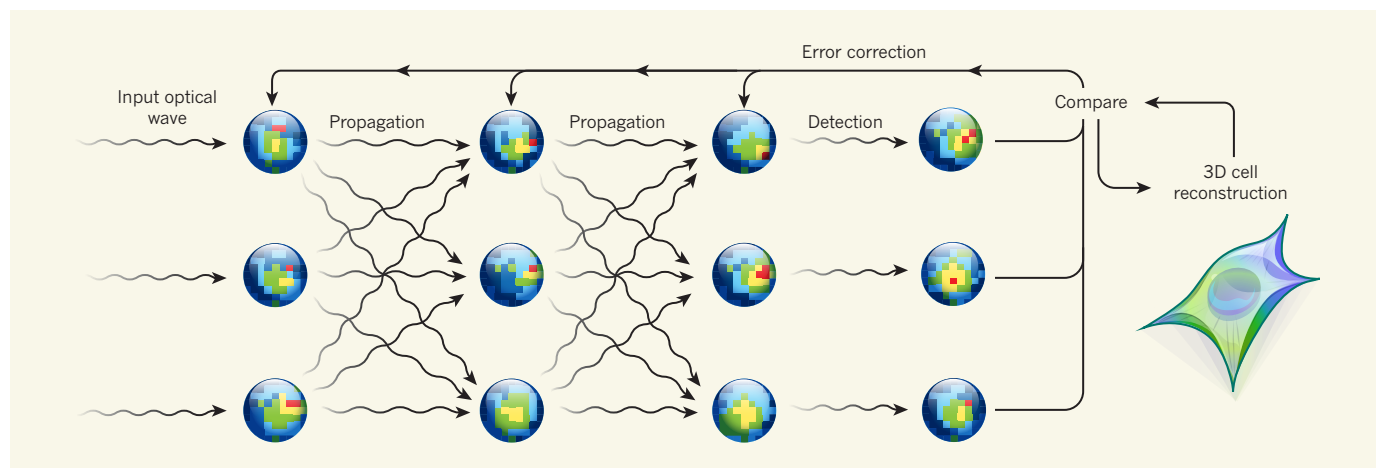


Figure 1 | 3D image reconstruction with artificial neural networks. Kamilov *et al.*¹ use an artificial neural network (ANN) algorithm to describe how the phase of optical light is modified as it propagates through a 3D biological sample (here, a cell). The sample is modelled as a series of layers. Each pixel (circles) of the 3D model corresponds to a node of the ANN. These are connected to

nodes in the subsequent layer (arrows) to represent the scattering of the input wavefront in the direction of propagation. The algorithm incorporates error correction by comparing a detector's measurements with the model's output — a 3D reconstruction of a cell's refractive index — and minimizing the difference between the two. (Figure adapted from Fig. 2 of the paper¹.)

computational imaging and artificial neural networks (ANNs). The latter underlie a popular machine-learning framework that has found many applications², from e-commerce and e-mail spam filtering to finding cat videos on YouTube. ANNs have been used to solve problems that involve big data (for example, image classification) and so they are a natural fit for computational microscopy.

Microscopists are swimming in data — they can easily collect terabytes of images in a few minutes. Easy access to large data sets creates the perfect opportunity for data-science approaches to image reconstruction. First, use all available knowledge about the sample (for example, an estimate of the number of bright spots within it) and about the imaging system (from optical physics) to constrain the problem, and then upload all the data to the computer and let the algorithm find the answer. Although there may not be an explicit analytical solution to the reconstruction problem using this approach, important information can still be teased out.

The authors use ANNs to attack the 3D phase-imaging problem, which is compounded by the complication of multiple scattering of light as it passes through a 3D biological sample. Multiple scattering is one of the most challenging problems in optics — if we solved it completely, we could see through fog, murky water or even human tissue. Physicists have tried for decades to undo scattering analytically, but it is difficult, if not impossible, to tackle large-scale problems that involve many scattering events. The authors' machine-learning approach is indirect (non-analytical), but gives a good solution that they verify experimentally.

Kamilov and co-workers adapt ANNs to work with the multi-slice method³, which has previously been used to describe multiple (dynamical) scattering of electrons in 3D crystal lattices. The authors model the target object as a set of slices: each slice is represented by a layer of the network and each pixel of the 3D object is represented by a network node (Fig. 1). The ANN's training data consist of a set of 2D holograms of the 3D object that are captured from different angles. The authors use a modified 'back-propagation' algorithm that predicts the 3D refractive index of the object by minimizing the differences between the training data and model solutions, with an added 'sparsity' constraint that enforces the smoothness of the solution. Multiple scattering is treated only in the general direction of the propagation — that is, backwards-reflected light is not included in the computations. Similar methods, applied to different hardware set-ups, have provided spatial resolution beyond the diffraction limit of an optical microscope⁴ or at the atomic scale in studies using electron microscopy⁵.

This work is part of a larger movement to revolutionize imaging techniques by

rethinking both the optical design and the post-processing of the images. Fully leveraging the power of machine learning for microscopy could lead to methods that can see inside the human body and resolve individual cells by overcoming multiple scattering. However, we are a long way off, and for this to be achieved, physicists and engineers need to account properly for complications arising from back-scattered light and for the directional dependence (anisotropy) of the objects' optical properties. In this quest, extremely large imaging data sets will surely be required and researchers may need to follow promising frontiers in data science (such as deep learning²), or invent new ones.

Kamilov and co-workers' shift away from analytical solutions allows them to find an answer to the 3D imaging reconstruction problem, but such an approach does not always have a provably correct solution. This is not a

problem in many of the applications of data science — no one dies if your cat-video search accidentally returns a dog video. But for scientific imaging applications, for example in medical settings, provability may be critical. As such, computational imaging brings a rich set of challenges for theorists and statisticians, as well as practitioners. ■

Laura Waller and Lei Tian are in the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, California 1770, USA. e-mail: waller@berkeley.edu

1. Kamilov, U. S. *et al. Optica* **2**, 517–522 (2015).
2. LeCun, Y., Bengio, Y. & Hinton, G. *Nature* **521**, 436–444 (2015).
3. Cowley, J. M. & Moodie, A. F. *Acta Crystallogr.* **10**, 609–619 (1957).
4. Tian, L. & Waller, L. *Optica* **2**, 104–111 (2015).
5. Van den Broek, W. & Koch, C. T. *Phys. Rev. Lett.* **109**, 245502 (2012).

ORGANIC CHEMISTRY

Natural polarity inverted

The concept of umpolung describes the reversal of the naturally occurring electrostatic polarization of chemical groups. It has now been used to make single mirror-image isomers of nitrogen-containing molecules. SEE LETTER P.445

FEDOR ROMANOV-MICHAILIDIS & TOMISLAV ROVIS

The action of biologically active molecules depends on the precise spatial arrangement of atoms that interact with biological targets. More than 95% of drug molecules¹ contain nitrogen atoms because they improve the cell permeability and water solubility of the compounds, and strengthen their interactions with biological targets. Methods for the spatially

selective assembly of nitrogen-containing molecules are therefore of considerable interest for drug discovery. Moreover, biological targets have a particular chirality (handedness). The ability to synthesize just one chiral form — one enantiomer — of biologically active compounds is thus also of great importance, because only molecules of the correct handedness will fit into their targets, in the same way that right-handed gloves best accommodate right hands. On page 445 of this issue, Deng and colleagues²

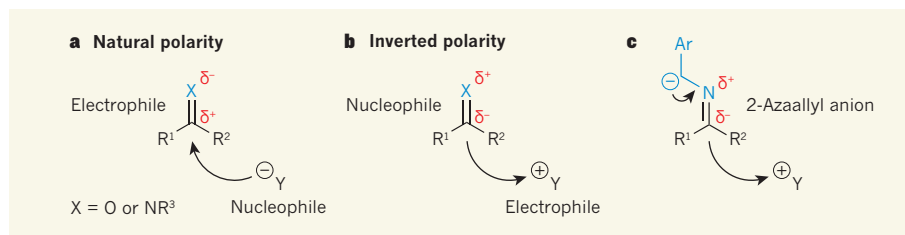


Figure 1 | Umpolung for ketones and imines. **a**, The natural polarization of ketones (in which $X = O$) and imines ($X = NR^3$) places partial positive charge (δ^+) on the carbon atom (and partial negative charge, δ^- , on X). These compounds are therefore electrophilic — attracted to areas of negative charge, such as those in nucleophilic molecules (Y^-). Ketones and imines are thus prone to attack by nucleophiles. R^1 to R^3 represent hydrocarbon groups; curly arrow represents electron movement. **b**, Umpolung describes the inversion of natural polarization in molecules. The application of umpolung to ketones and imines would make them nucleophilic, and prone to attacking electrophiles. **c**, Deng and colleagues² report that 2-azaallyl anions act as umpolung forms of imines. Ar represents a 4-nitrophenyl group.

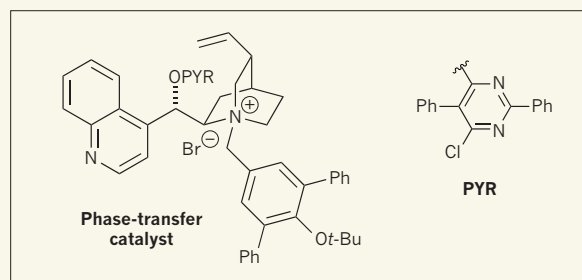
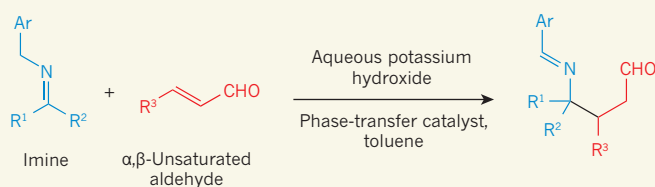


Figure 2 | Catalytic enantioselective reactions of umpolung imines. Deng and colleagues² report a carbon–carbon bond-forming reaction between imines and α,β -unsaturated aldehydes. The reaction depends on an umpolung form of the imine, and occurs in the presence of a phase-transfer catalyst. The catalyst transfers a base (potassium

hydroxide) from an aqueous solution to toluene (the solvent in which the reaction occurs), and also induces the product to form predominantly as a single enantiomer (mirror-image isomer). R^1 to R^3 represent hydrocarbon groups; Ar represents a 4-nitrophenyl group; Ph, phenyl; *t*-Bu, tertiary butyl ($C(CH_3)_3$).

report a method that solves both problems by reversing the natural electrostatic polarization of groups called imines.

Every chemical group of a given molecule is characterized by a pattern of electrostatic polarization that dictates the group's reactivity towards other molecules. The polarization of carbonyl ($C=O$) and imine ($C=N$) groups places partial positive charge at the carbon atom of the group, making it electrophilic — attracted to areas of negative charge (Fig. 1a). Molecules that bear partial negative charge are called nucleophiles, and tend to attack electrophilic carbon atoms, thereby creating a chemical bond.

The 'natural' polarization of a group can sometimes be reversed, so that electrophilic sites become nucleophilic and vice versa. This concept is known as umpolung³, from the German term for reversal of polarity. The umpolung of an imine or of a carbonyl-containing compound, such as a ketone, would place partial negative charge at the carbon atom, rendering the atom nucleophilic (Fig. 1b). The development of synthetic strategies based on umpolung opens up fresh vistas for the construction of biologically active molecules.

Several catalytic strategies^{4–6} have been designed to invert the natural reactivity pattern of carbonyl-containing compounds. An analogous catalytic strategy that allows the enantioselective synthesis of nitrogen-containing compounds from imines is highly desirable for drug-discovery research. Deng and co-workers' ingenious solution to this problem relies on the reaction of transiently formed molecules called 2-azaallyl anions (Fig. 1c) with carbon-based electrophiles.

Their reaction builds on a widely used concept⁷ first pioneered by the chemist Vadim Soloshonok and subsequently improved by Deng's research group⁸ and by the chemist Yian Shi and his group⁹: the use of one enantiomer of a base to isomerize imines to form enantiomerically enriched amine compounds. The proposed intermediate 2-azaallyl anions are similar in reactivity to hydrazone compounds⁶ that have been used in umpolung

reactions of carbonyl-containing compounds, and behave as carbon nucleophiles.

Several non-enantioselective transformations have previously been reported^{10–13}, and the sole example of a highly enantioselective carbon–carbon ($C-C$) bond-forming reaction with 2-azaallyl anions was a palladium-catalysed coupling with carbon electrophiles¹⁴.

“The reaction products are modified versions of imines, and can be readily converted into a variety of other nitrogen-containing compounds.”

and also inducing enantioselectivity. The reaction products are modified versions of imines (Fig. 2), and can be readily converted into a variety of other nitrogen-containing compounds.

The most interesting aspect of this work is the clever catalyst design (Fig. 2). It was developed from a quinine compound that was originally derived from *Cinchona* plants and which has previously been used as a scaffold for other phase-transfer catalysts¹⁵. The authors found that the prototypical catalyst delivered a different product to the one they were targeting, but by manipulating the catalyst's groups they were able to redirect the course of the reaction. A large, electron-rich group on the catalyst's nitrogen atom was required for high reactivity and enantioselectivity.

In the previous work by Deng's group⁸, only highly activated imines could be used in the reaction, but the new catalyst allows a wide variety of imines to participate with nearly equal facility. Furthermore, the reaction proceeds with remarkable enantioselectivity, and yields the amine products with high fidelity. It is also easy to set up and tolerates air and

moisture from the atmosphere. It remains to be seen whether the substrate scope can be further extended by manipulating the catalyst's structure, to allow the use of simpler imines and less reactive electrophiles than those reported in this paper. It should also be noted that the catalyst is not currently commercially available, but it seems to be uncomplicated to synthesize.

Deng and co-workers' findings illustrate the power of catalyst development for organic synthesis, and provide a straightforward route to chiral amines. Their method also adds to the arsenal of established umpolung strategies for carbonyl derivatives, and is complementary to other such methods. ■

Fedor Romanov-Michailidis and Tomislav Rovis are in the Chemistry Department, Colorado State University, Fort Collins, Colorado 80523, USA.
e-mail: tomislav.rovis@colostate.edu

- Vitaku, E., Smith, D. T. & Njardarson, J. T. *J. Med. Chem.* **57**, 10257–10274 (2014).
- Wu, Y., Hu, L. & Deng, L. *Nature* **523**, 445–450 (2015).
- Seebach, D. *Angew. Chem. Int. Ed.* **18**, 239–258 (1979).
- Denmark, S. E. & Wilson, T. W. *Nature Chem.* **2**, 937–943 (2010).
- Flanigan, D. M., Romanov-Michailidis, F., White, N. A. & Rovis, T. *Chem. Rev.* <http://dx.doi.org/10.1021/acs.chemrev.5b00060> (2015).
- Brehme, R., Enders, D., Fernandez, R. & Lassalella, J. M. *Eur. J. Org. Chem.* **2007**, 5629–5660 (2007).
- Han, J., Sorochinsky, A. E., Ono, T. & Soloshonok, V. A. *Curr. Org. Synth.* **8**, 281–294 (2011).
- Wu, Y. & Deng, L. *J. Am. Chem. Soc.* **134**, 14334–14337 (2012).
- Xiao, X., Xie, Y., Su, C., Liu, M. & Shi, Y. *J. Am. Chem. Soc.* **133**, 12914–12917 (2011).
- Matsumoto, M., Harada, M., Yamashita, Y. & Kobayashi, S. *Chem. Commun.* **50**, 13041–13044 (2014).
- Liu, X., Gao, A., Ding, L., Xu, J. & Zhao, B. *Org. Lett.* **16**, 2118–2121 (2014).
- Ogle, J. W., Zhang, J., Reibenspies, J. H., Abboud, K. A. & Miller, S. A. *Org. Lett.* **10**, 3677–3680 (2008).
- Kim, H. *et al.* *J. Am. Chem. Soc.* **130**, 12184–12191 (2008).
- Zhu, Y. & Buchwald, S. L. *J. Am. Chem. Soc.* **136**, 4500–4503 (2014).
- Shirakawa, S. & Maruoka, K. *Angew. Chem. Int. Ed.* **52**, 4312–4348 (2013).

Speed cells in the medial entorhinal cortex

Emilio Kropff^{1,2}, James E. Carmichael^{1†}, May-Britt Moser¹ & Edvard I. Moser¹

Grid cells in the medial entorhinal cortex have spatial firing fields that repeat periodically in a hexagonal pattern. When animals move, activity is translated between grid cells in accordance with the animal's displacement in the environment. For this translation to occur, grid cells must have continuous access to information about instantaneous running speed. However, a powerful entorhinal speed signal has not been identified. Here we show that running speed is represented in the firing rate of a ubiquitous but functionally dedicated population of entorhinal neurons distinct from other cell populations of the local circuit, such as grid, head-direction and border cells. These 'speed cells' are characterized by a context-invariant positive, linear response to running speed, and share with grid cells a prospective bias of ~50–80 ms. Our observations point to speed cells as a key component of the dynamic representation of self-location in the medial entorhinal cortex.

Grid cells in the medial entorhinal cortex (MEC) are unique in their spatial code^{1,2}. Unlike other place-modulated neurons, their population firing pattern not only repeats periodically within a given environment¹, but also seems to apply equally to all explored environments³, reflecting the uniformity of space despite the unevenness of contextual details. This property makes grid cells ideal candidates for a path integration-based representation of space^{1,4–11}. In such a scheme, running speed is integrated across short time windows to obtain the instantaneous displacement of the animal, which, in conjunction with head-direction input, is used to update the representation of the animal's position. Any path integration mechanism thus requires running speed as a major input. However, while speed correlates marginally with entorhinal theta frequency¹² and the firing rate of some grid cells^{13,14}, the existence and nature of a reliable and locally available speed signal has remained unclear. The aim of the present study was to determine whether speed is represented in separate classes of MEC cells.

Experimental control of running speed

We began by recording neuronal activity under strict experimental control of the animal's running speed. Rats traversed a 4 m long linear track with their body confined inside a computer-driven bottomless frame that was moved along the track at a pre-set speed (an experimental car similar in concept to a 'Flintstones' car'; Fig. 1a). Since the car had no floor, it compelled the animal to engage in natural locomotion at the experimenter-determined speed in order to reach the end of the track, where a food reward was delivered. During running, cells were recorded across all layers of the MEC (Extended Data Fig. 1).

In initial experiments, we either trained rats to run fast on one half of the track and slow on the other, with a sharp transition in the middle (382 cells, three rats), or speed was increased proportionally to the distance from one of the track ends (282 cells, two rats). While spatial maps were not disrupted (Extended Data Fig. 2), the firing rate

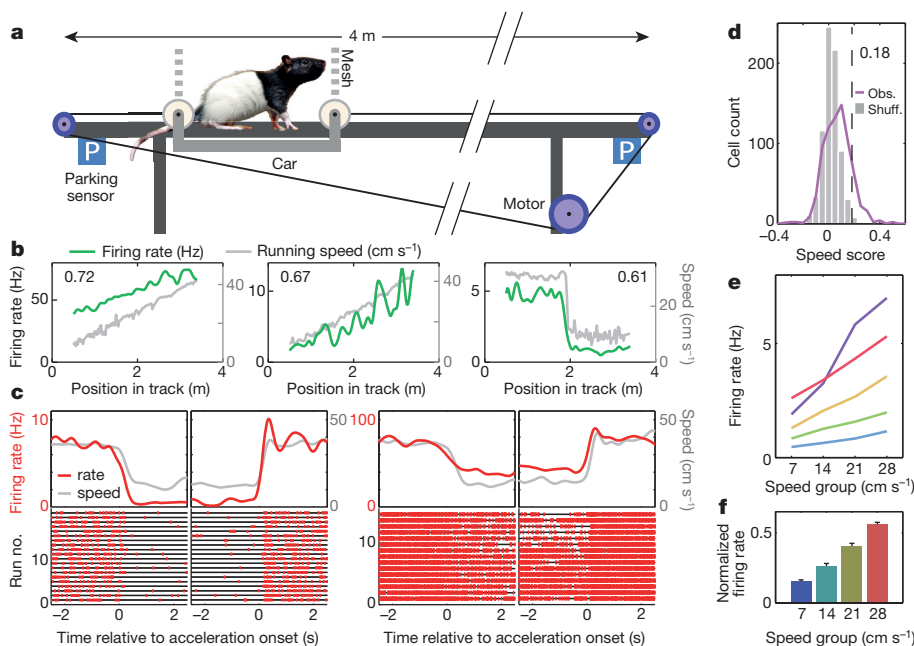


Figure 1 | Speed-responsive MEC cells in a linear task. **a**, Bottomless car. **b**, Mean firing rate (green) and running speed (grey) as a function of position for three representative speed cells in the MEC. Left and middle, linear speed protocol; right, step protocol. Pearson correlations between instantaneous running speed and firing rate are indicated. **c**, Representative speed cells during decelerating and accelerating events of the step protocol (left and right subpanels, respectively). Top, firing rate (red, left axis) and running speed (grey, right axis) as a function of time relative to the event onset. Bottom, spike raster plots. **d**, Distributions of observed speed scores (purple curve, 'Obs.') and 100 shuffles per cell (grey bars, 'Shuff.'). Counts normalized by number of shuffles. Dashed line shows the 99th percentile of the shuffled distribution (0.18). **e**, Tuning curves of five representative speed cells in **d**. **f**, Normalized average firing rates (see Methods) for all 98 speed cells in all four speed groups (means \pm s.e.m.).

¹Kavli Institute for Systems Neuroscience and Centre for Neural Computation, Norwegian University of Science and Technology, Olav Kyrres gate 9, MTF5, 7491 Trondheim, Norway. ²Leloir Institute, IIBBA - CONICET, Buenos Aires, C1405BWE, Argentina. [†]Present address: Department of Psychological and Brain Sciences, Dartmouth College, Hanover, New Hampshire 03755, USA.

of some cells recorded in these protocols followed the speed profile (Fig. 1b), with fast transitions in firing rate at each change in running speed (Fig. 1c).

In order to disentangle running speed from the position of the animal, the same track segments were traversed, in separate experiments, at constant running speeds that alternated randomly from run to run between 7, 14, 21 and 28 cm s^{-1} . The majority of data in the bottomless car task (754 cells from ten rats) were collected with this four-speed protocol. To identify speed-responsive neurons, we calculated a speed score for each cell, defined as the Pearson product-moment correlation between instantaneous firing rate and running speed, on a scale from -1 to 1 . Cells with speed scores higher than the 99th percentile of a shuffled distribution (a value of 0.18) were classified as speed cells. A total of 98 MEC cells (13%) passed this criterion (Fig. 1d, e), significantly more than expected by chance (expected, 7.5 cells; $P = 10^{-74}$). As a population, these cells showed significant differences in normalized firing rate between all four blocks of constant running speed (Fig. 1f; Kruskal–Wallis and Tukey–Kramer tests, $P < 0.01$). The slope and y intercept of linear regression lines for firing rate as a function of speed were highly dispersed across cells, with frequent cases of non-zero firing at very low speed (Extended Data Fig. 3a; see examples in Fig. 1b, c, e). Negative, linear responses to speed were also observed, although only in a marginal sub-population of 16 cells (2%) that was not analysed further (Extended Data Fig. 3b–d).

Speed cells during free foraging

Analyses in the bottomless car do not address directly the question of how much overlap there is between the speed-cell population and other entorhinal cell types, which have two-dimensional firing patterns. Thus, we performed classical free-foraging recordings in a 1 m wide square box, where 17 rats covered a wide range of instantaneous speeds, typically from 0 to 50 cm s^{-1} . We recorded 2,497 MEC cells and obtained for each cell a speed score and a rate-by-speed tuning

curve using 2- cm s^{-1} bins from 0 to 50 cm s^{-1} . Again, many cells had firing rates that increased linearly with speed (Fig. 2a and Extended Data Figs 4a and 5). Instantaneous firing rate and running speed exhibited considerable co-variation (Fig. 2b). In behaviourally unfiltered data, as many as 51% of the neurons had a speed score that passed the classification threshold determined by the 99th percentile of a distribution of shuffled data (Fig. 2c). This large proportion might reflect a difference in the network state between rest and active navigation rather than a genuine correlation between running speed and firing rate, a problem that was not present in the bottomless car, where resting periods were left out of the analysis. To overcome this issue, we redefined the speed score by filtering out static periods (speed $< 2 \text{ cm s}^{-1}$). With this stricter criterion, used in all subsequent open-field analyses, the threshold defined by the 99th percentile of the shuffled distribution (0.18) was passed by 385 neurons, or 15% of all MEC cells (Fig. 2d and Extended Data Fig. 6f)—a percentage almost identical to the estimate from the bottomless car. Slopes and y intercepts of regression lines for firing rate as a function of speed were as dispersed as in the car (Extended Data Fig. 3a). Cells with similar properties were found in the hippocampus (Fig. 2e and Extended Data Fig. 7b). Out of 964 hippocampal neurons that were active in the open field, 96 cells (10%) passed the threshold determined by the 99th percentile of the shuffled distribution (0.19) (Fig. 2f).

Once we had established that speed modulation is similar in the open field and the bottomless car, we could investigate whether MEC speed cells form a population of their own. We compared their properties with those of grid, head-direction and border cells, classified respectively by their gridness score^{13,15}, mean vector length¹⁶, and border score^{17,18}, with thresholds obtained by shuffling of spike times (Extended Data Fig. 6a). Out of 2,497 cells, 518 (21%) were classified as grid cells, 398 (16%) as head-direction cells, and 99 (4%) as border cells (Extended Data Fig. 6c). The intersection between speed cells and any of these cell populations was small and, for grid cells and

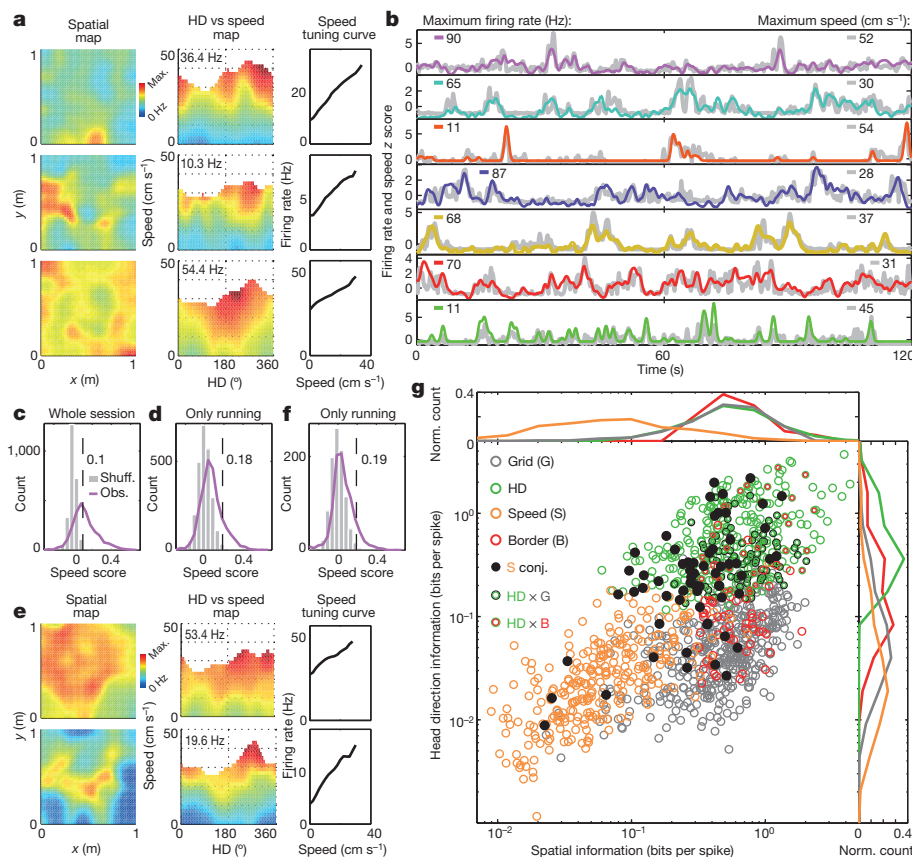


Figure 2 | Entorhinal speed cells form a population of their own. **a**, Firing rate as a function of position, head direction (HD), and running speed in three representative MEC speed cells during free foraging in an open field. Each row shows one cell. Left, colour-coded spatial rate maps. Scale bar to the right. Middle, firing rate (colour) as a function of head direction (x axis) and running speed (y axis). Right, firing rate as a function of running speed. **b**, Traces showing z scores for firing rate (colour) and speed (grey) for seven representative entorhinal speed cells during 2 min of free foraging (different sessions). Maximum values of firing rate and speed are indicated (left and right, respectively). **c**, Distributions of observed speed scores (purple curve, 'Obs.') and 100 shuffles per cell (grey bars, 'Shuff'; counts normalized by number of shuffles) for all MEC cells in the open field. Dashed line, 99th percentile of the shuffled distribution (value at the top). **d**, As **c**, but including only running periods (speed $> 2 \text{ cm s}^{-1}$). **e**, Speed cells in the hippocampus (as in **a**). **f**, Distribution of hippocampal speed scores, excluding static periods as in **d**. **g**, Logarithmic-scale scatter plot showing head-direction and spatial information of 1,206 neurons (dots) classified by statistical criteria as speed, grid, head-direction or border cells. Cell identity is colour coded. Black dots represent 62 conjunctive speed cells (S conj.), out of 385, that satisfied criteria for at least one additional cell type. Curves show normalized counts of spatial (top) and head-directional (right) information for each cell category.

head-direction cells, significantly lower than expected by chance. Only 16 speed cells met the criterion for grid cells (expected, 79.9; $P = 10^{-17}$), 42 met the head-direction criterion (expected, 61.4; $P = 0.005$), and 11 met the criteria for border cells (expected, 15.3; $P = 0.17$) (Extended Data Figs 4b, 6c, d and 7a). These numbers contrast with the overlap between spatial and directional cells, which was never below chance (Extended Data Fig. 6c). Almost half of the speed cells with a dual classification had low in-field speed scores (speed scores restricted to the data inside the spatial or directional fields), suggesting that, in these cells, the speed modulation was indirect, caused by interactions between speed and other behavioural variables (Extended Data Fig. 6h). Similarly, the amplitude of grid fields in the bottomless car was not significantly modulated by speed (Extended Data Fig. 3e). Consistent with this functional separation of the speed cells, we found that they had a population distribution of spatial and head-direction information per spike around one order of magnitude below that of grid, head-direction and border cells (Fig. 2g and Extended Data Fig. 6d,i). The latter cell types had similar distributions of speed score (Extended Data Fig. 6g), typically lower than the criterion for speed cells and not very different from the distribution of shuffled data with a 2 cm s^{-1} threshold (Fig. 2d). In sum, because of their distinct firing characteristics and the low levels of overlap with other cell types, MEC speed cells seem to form a population of their own. This conclusion does not apply to hippocampal data, where in-field speed scores were often higher than average scores (Extended Data Fig. 6c, e, h) and place cells exhibited a low but significant modulation by speed in the bottomless car (Extended Data Fig. 3e).

Many speed cells had properties of fast-spiking cells, classified as neurons with a mean firing rate above 10 Hz (Extended Data Fig. 6g) and a spike width below 0.3 ms (ref. 19). Of 385 speed cells, 95 (25%) were classified as fast-spiking in the MEC, and 27 out of 96 (28%) in the hippocampus. Among 1,178 grid, head-direction, border and place cells, only four (0.3%) passed the criteria for fast-spiking cells. Speed cells were present in all MEC layers, with a rather homogeneous distribution (minimum 14%; maximum 18%; Extended Data Fig. 6f).

The speed code is context-invariant

We next asked whether the speed code expressed in the population of MEC speed cells could be used for path integration. For this to happen, not only should it be possible to decode running speed from the activity of speed cells, but the decoding should be context-invariant. We analysed data from previous experiments^{3,13} in which spike activity was recorded from the MEC in two rooms (A and B; sequence AB followed by a second session in A named A'; eight rats). Twenty speed cells were identified. As expected³, grid cells fired at different locations relative to the box walls in the two rooms (Fig. 3a). Speed cells, in contrast, had invariant speed scores and tuning curves (Fig. 3b, c and Extended Data Fig. 8b, c). In one case, a rat had four simultaneously recorded speed cells, a situation fit for decoding. Two simple linear decoders, trained with data from these four cells in either A or B, were tested in A'. Reconstructed (decoded) speed was highly correlated with tracked speed, irrespective of whether room A or room B was used to train the decoder (A, $r = 0.75$; B, $r = 0.74$) (Fig. 3d). In general, the match between reconstructed and actual speed increased with the

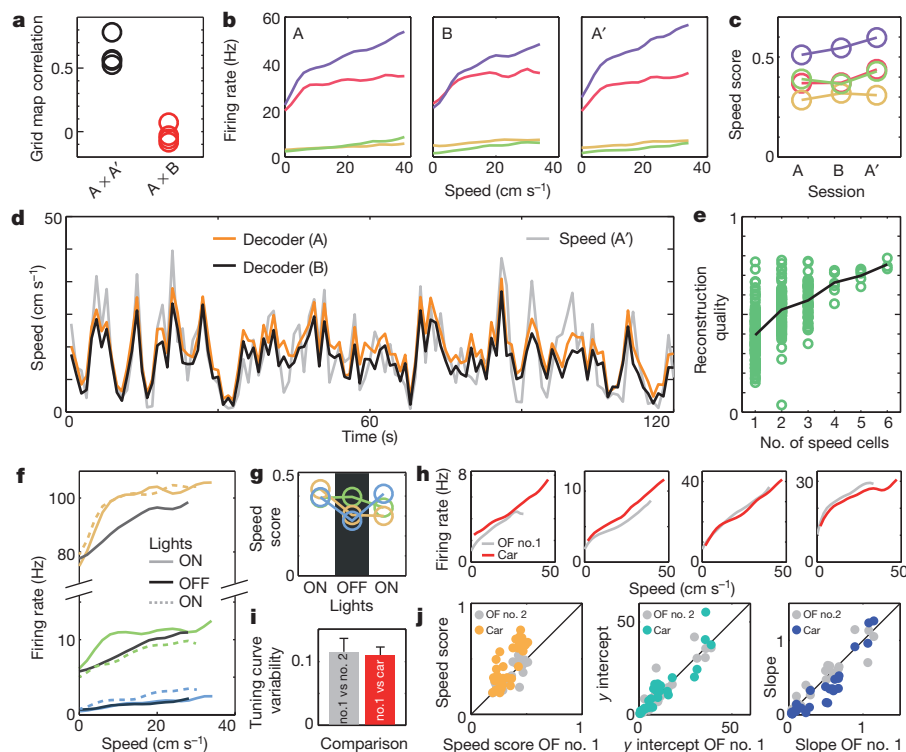


Figure 3 | Invariance of the entorhinal speed code. **a**, Correlation of grid maps on two trials in the same room (A versus A') or in different rooms (A versus B) in a representative rat. **b**, **c**, Tuning curves (**b**) and speed scores (**c**) of four simultaneously recorded speed cells (unique colours) on successive trials in rooms A and B. Note room-independent speed–rate relationships. **d**, Two linear decoders, trained with the activity of these four speed cells in either A or B, were used to decode running speed in A'. Reconstructions from A and B are very similar (Pearson correlation = 0.99) and match actual speed in A' (reconstruction quality, or Pearson correlation with running speed, A, 0.75; B, 0.74). **e**, Reconstruction quality as a function of the number of simultaneously recorded speed cells (all trials in open field). The training data

set comprised the initial 70% of the session and the decoders were tested on the remaining 30%. **f**, **g**, Tuning curves (**f**) and speed scores (**g**) of three simultaneously recorded speed cells during two regular sessions with a trial in darkness in-between. **h**, Tuning curves in open-field (OF) and bottomless car trials for speed cells from rat 14566, which was trained to run twice as fast in the car as in the open field (Extended Data Fig. 8f). Note the gain invariance. **i**, Tuning curves had similar variability across open-field sessions and between open-field and car session (Mann–Whitney U -test $P = 0.34$; all speed cells from rat 14566). **j**, Speed score, y intercept and slope for second open-field trial (grey) or car trial (colour) against first open-field trial (same cells as **i**).

number of simultaneously recorded cells, reaching an average Pearson correlation of ~ 0.75 for six cells (Fig. 3e; 385 speed cells; all open-field sessions where at least one speed cell was recorded). In an experiment where three speed cells were recorded with room lights on and off in an on-off-on sequence¹, the speed code was similarly invariant (Fig. 3f, g and Extended Data Fig. 8d), suggesting that optic flow is dispensable. Finally, the speed code was also largely invariant across experimental tasks, as demonstrated when speed cells were recorded in two open-field sessions with a bottomless car session in-between (Extended Data Fig. 8e). The tuning curves were similar, with no slope adaptation, even in a rat that showed a twofold difference in average speed between open-field and bottomless car trials (Fig. 3h–j and Extended Data Fig. 8f). In sum, MEC speed cells express a context-invariant speed code that can be used to decode actual running speed across a variety of experimental manipulations.

The entorhinal speed code is prospective

We asked if speed cells represent instantaneous speed or have retrospective or prospective components, which have been reported for place cells and grid cells under a variety of circumstances^{20–24}. If place and grid cells are driven by path integration based on input provided by speed cells, the temporal bias might be derived from the speed signals themselves. To test this hypothesis, we calculated correlations between running speed and different temporal shifts of the instantaneous firing rate, in order to find the shift that maximized the correlation. The firing rate of the MEC speed cells correlated better with future speed than simultaneous or past speed, both in bottomless car trials (all trials pooled) and in the open field (Fig. 4a, b; correlation maxima at time shifts of 54–82 ms; $P < 0.01$; Extended Data Fig. 9b). This bias was present only in theta-modulated cells (37% of all speed cells), where the speed-related firing ramped up in

a characteristic pattern during the course of the theta cycle (Extended Data Fig. 9).

We next examined the consequences of prospective path integration on the firing locations of grid cells. The expected amount of anticipation in a prospective path integrator can be estimated directly for episodes of constant running speed (see Methods, equation (2)). If the speed signal anticipates running speed by a fixed time interval τ , the resulting spatial anticipation is proportional to the running speed, with τ as the coefficient of proportionality. Using only constant running episodes from the four-speed experiments, we compared the position of the same entorhinal firing fields (putative grid fields) at 7, 14, 21 and 28 cm s^{-1} (Extended Data Fig. 3f). The average field was linearly shifted to earlier positions for higher speeds, with a slope similar to the temporal anticipation of MEC speed cells ($\tau = 80$ ms, $r = 0.97$) (Fig. 4b). This is compatible with the idea that grid cells are driven by path integration based on input from speed cells. A similar link was not observed in the hippocampus, where speed cells showed a significant retrospective effect (τ between -89 ms and -59 ms) and place cells showed no temporal bias ($\tau = 1$ ms, $r = -0.07$) (Fig. 4a, b and Extended Data Fig. 9b).

The magnitude of the anticipatory shift of the grid fields increased during acceleration episodes. Although we did not find a significant population of cells directly tuned by acceleration (Extended Data Fig. 6b), in the MEC the firing rate of the speed cells was positively modulated by acceleration, as expected from their prospective nature (see Methods, equation (1)) (Fig. 4c, top; absolute threshold: 50 cm s^{-2} ; Friedman's test for acceleration effects in both tasks, $P < 0.01$). This increase in the firing of speed cells would make the path integrator run faster and thus generate a larger spatial anticipation. To test this idea, we estimated the impact of positive versus negative acceleration, filtered at an absolute threshold of 50 cm s^{-2} ,

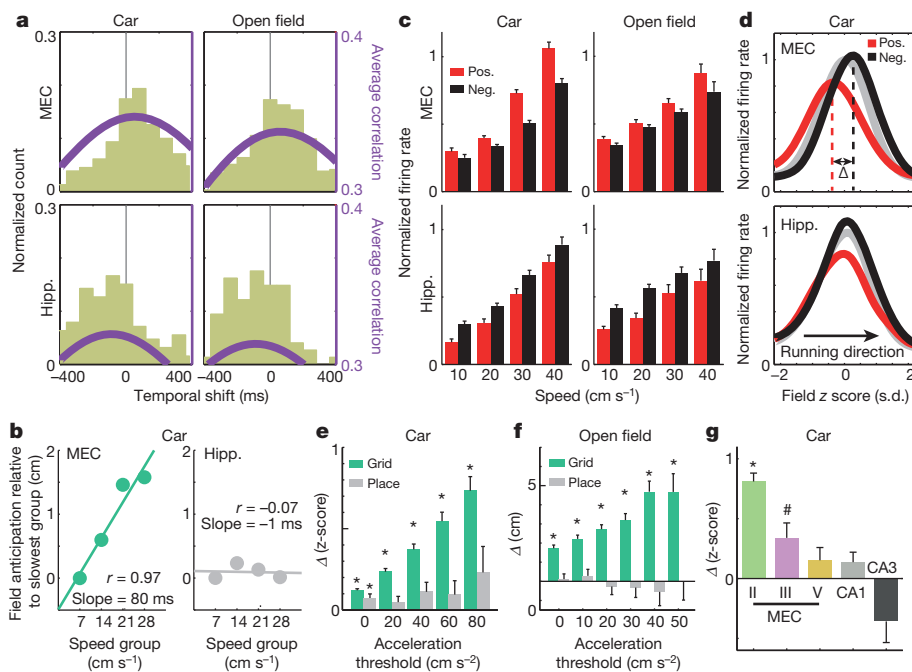


Figure 4 | Prospective coding of entorhinal speed cells and grid cells.

a, Correlation between running speed and temporal shifts of instantaneous firing rate for speed cells in MEC (top) and hippocampus (hipp.; bottom). Left, car; right, open field. Green bars show normalized counts of temporal shifts that maximize correlation. Average correlation curves are shown in purple. Note prospective bias in MEC speed cells and retrospective bias in hippocampal cells. **b**, Relation between speed and spatial anticipation (peak field position relative to that of 7 cm s^{-1} group). Data from four-speed car experiment. Note linear relation in grid cells (MEC) but not place cells (hippocampus). **c**, Normalized activity of speed cells (mean \pm s.e.m.) as a function of speed during intervals of

positive (red) and negative (black) acceleration (absolute threshold, 50 cm s^{-2}). **d**, Average fields of putative grid cells (top) and place cells (bottom) in the car after filtering for extreme acceleration (as in c) or without filtering (grey). Δ is the difference in spatial anticipation during positive compared to negative acceleration. **e**, Spatial shifts (Δ) corresponding to different acceleration thresholds (mean \pm s.e.m.) for grid cells (green) and place cells (grey) ($*P < 0.01$). **f**, Spatial shifts as in e, but for open-field sessions (see Methods) ($*P < 0.01$). **g**, Spatial shifts (Δ) for spatially modulated cells in the bottomless car classified by recording location and MEC layer (absolute acceleration threshold, 50 cm s^{-2} ; mean \pm s.e.m.; $*P < 0.01$, $\#P < 0.05$).

on the average position of grid fields in all bottomless car trials. Again, we found an anticipatory shift in field position in MEC cells due to positive acceleration (Fig. 4d and Extended Data Fig. 10a–e). The anticipatory shift increased with the absolute acceleration threshold (Fig. 4e, Mann–Whitney *U*-test for grid cells after Holms–Bonferroni correction, $P < 0.01$). A similar effect was found for grid cells in the open field (Fig. 4f, $P < 0.01$). The shift was significant only in MEC layer II, where it was large (Wilcoxon signed-rank test after Holms–Bonferroni correction, $P < 0.01$), and in layer III, where it was small ($P < 0.05$) (bottomless car data, Fig. 4g and Extended Data Fig. 10a). It was strongly modulated by theta activity (Extended Data Fig. 10f–h), in agreement with models suggesting that path integration takes place on a theta-cycle basis⁷. In the hippocampus, speed cells showed significant negative modulation by acceleration, compatible with retrospective coding (equation (1)), but place cells exhibited no significant spatial shift (Fig. 4c–g and Extended Data Fig. 9b; Friedman's test, $P < 0.01$). Taken together, these observations support the idea that entorhinal speed cells contribute to the firing of grid cells via path integration, a process that does not seem to take place in the hippocampus.

Discussion

The main finding of our study is the discovery of a functionally dedicated population of speed cells in the MEC. These cells, which represent a considerable fraction of the MEC neurons (~15% across all layers), are characterized by a positive, linear response to running speed, and low levels of spatial and directional information. The speed response was independent of visual input, consistent with the idea that the signal is at least partly derived from proprioceptive or motor-efference information in the mesencephalon²⁵. Neurons with similar characteristics were found in the hippocampus (~10%). The observations are in agreement with prior anecdotal reports of a speed-modulated axon in the hippocampus²⁶ and one in or around the presubiculum²⁷. The presence of speed-modulated place cells is also consistent with earlier work^{28,29}. Unlike the hippocampal cells, however, the speed-cell population in the MEC exhibited little overlap with other cell types. Earlier work has demonstrated correlations between running speed and firing rate in grid cells^{13,14}, but the present data shows that when speed is experimentally disentangled from space, acceleration and behavioural state, the grid-cell population exhibits no speed modulation, and only around 1% of all grid cells show a robust speed response. Reconstruction of instantaneous speed would thus be possible only with input from hundreds or thousands of grid or head-direction cells. Speed cells, in contrast, allow for accurate decoding of speed from the activity of only 4–6 specialized cells. With complex functional properties such as prospectiveness and a unique modulation by theta phase, speed cells can hardly be thought of merely as passive integrators of the diffuse speed information coded by other MEC populations.

The existence of speed-responsive cells in the entorhinal-hippocampal network has implications for how spatial maps are updated as animals navigate through an environment. Path-integration-dependent models make use of a speed signal, coded either in the frequency of membrane oscillations^{9–11,30,31} or in the firing rate of dedicated neurons^{4–8}. The speed signal is used to dynamically update grid-cell activity in accordance with the animal's movement in space. Two important requirements must be met for the speed signal to enable efficient path integration. The first is a linear speed–rate relationship, which makes the temporal integration of the signal proportional to the displacement of the animal, allowing for a simple combination of multiple inputs to the same cell. The second requirement is contextual invariance. Our data show that the speed code is linear and invariant across environments, in darkness and in light, with no gain adaptation for different behavioural conditions. The universality of the speed code has a great advantage in that the path integrator needs to be trained only once in the animal's lifetime, allowing it to be used effectively in novel environments and in the absence of strong contextual cues, precisely where it is most needed.

Finally, MEC speed cells and grid cells are linked by a common prospective bias of ~50–80 ms, with a strong theta modulation suggesting that path integration occurs on a theta-cycle basis. The temporal bias is consistent with previous reports of alternating modes of prospective and retrospective firing in grid cells²³. However, the present observations suggest that these modes reflect positive- and negative-acceleration episodes, respectively. The spatial shift is purely prospective with respect to an unbiased spatial reference, such as the one provided by low speed rather than the total average of the data. Positive acceleration at the beginning of a movement may put the position estimated by the grid network ahead of the actual one, while negative acceleration at the end of a movement might compensate, bringing estimated and actual positions back together when the animal stops. In contrast to the observations in the MEC, no direct link could be established between speed cells and place cells in the hippocampus. Place cells may under some circumstances inherit prospective firing from grid cells²², but the present data suggest that, in general, temporal biases in the hippocampus follow a logic of their own, independent of path-integration processes that take place in the MEC.

The unique association of the predictive code with layer II grid cells is among the major functional differences described so far in the MEC circuit, and might therefore provide a key to understanding the computational steps underlying the dynamic representation of space. How the prospective speed signal is generated, why it is translated primarily to grid cells of layer II, how theta oscillations contribute to this process, and how the prospective firing in layer II interacts with non-prospective activity in other parts of the network are questions that remain to be addressed.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 March; accepted 3 June 2015.

Published online 15 July 2015.

- Hafting, T., Fyhn, M., Molden, S., Moser, M. B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806 (2005).
- Moser, E. I. et al. Grid cells and cortical representation. *Nature Rev. Neurosci.* **15**, 466–481 (2014).
- Fyhn, M., Hafting, T., Treves, A., Moser, M. B. & Moser, E. I. Hippocampal remapping and grid realignment in entorhinal cortex. *Nature* **446**, 190–194 (2007).
- McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I. & Moser, M. B. Path integration and the neural basis of the 'cognitive map'. *Nature Rev. Neurosci.* **7**, 663–678 (2006).
- Fuhs, M. C. & Touretzky, D. S. A spin glass model of path integration in rat medial entorhinal cortex. *J. Neurosci.* **26**, 4266–4276 (2006).
- Burak, Y. & Fiete, I. R. Accurate path integration in continuous attractor network models of grid cells. *PLOS Comput. Biol.* **5**, e1000291 (2009).
- Navratilova, Z., Giocomo, L. M., Fellous, J. M., Hasselmo, M. E. & McNaughton, B. L. Phase precession and variable spatial scaling in a periodic attractor map model of medial entorhinal grid cells with realistic after-spike dynamics. *Hippocampus* **22**, 772–789 (2012).
- Couey, J. J. et al. Recurrent inhibitory circuitry as a mechanism for grid formation. *Nature Neurosci.* **16**, 318–324 (2013).
- Burgess, N., Barry, C. & O'Keefe, J. An oscillatory interference model of grid cell firing. *Hippocampus* **17**, 801–812 (2007).
- Hasselmo, M. E. & Brandon, M. P. A model combining oscillations and attractor dynamics for generation of grid cell firing. *Front. Neural Circuits* **6**, 30 (2012).
- Bush, D. & Burgess, N. A hybrid oscillatory interference/continuous attractor network model of grid cell firing. *J. Neurosci.* **34**, 5065–5079 (2014).
- Jeewajee, A., Barry, C., O'Keefe, J. & Burgess, N. Grid cells and theta as oscillatory interference: electrophysiological data from freely moving rats. *Hippocampus* **18**, 1175–1185 (2008).
- Sargolini, F. et al. Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science* **312**, 758–762 (2006).
- Wills, T. J., Barry, C. & Cacucci, F. The abrupt development of adult-like grid cell firing in the medial entorhinal cortex. *Front. Neural Circuits* **6**, 21 (2012).
- Stensola, H. et al. The entorhinal grid map is discretized. *Nature* **492**, 72–78 (2012).
- Langston, R. F. et al. Development of the spatial representation system in the rat. *Science* **328**, 1576–1580 (2010).
- Solstad, T., Boccara, C. N., Kropff, E., Moser, M. B. & Moser, E. I. Representation of geometric borders in the entorhinal cortex. *Science* **322**, 1865–1868 (2008).
- Bjerknes, T. L., Moser, E. I. & Moser, M. B. Representation of geometric borders in the developing rat. *Neuron* **82**, 71–78 (2014).

19. Buetfering, C., Allen, K. & Monyer, H. Parvalbumin interneurons provide grid cell-driven recurrent inhibition in the medial entorhinal cortex. *Nature Neurosci.* **17**, 710–718 (2014).
20. Muller, R. U. & Kubie, J. L. The firing of hippocampal place cells predicts the future position of freely moving rats. *J. Neurosci.* **9**, 4101–4110 (1989).
21. Ferbinteanu, J. & Shapiro, M. L. Prospective and retrospective memory coding in the hippocampus. *Neuron* **40**, 1227–1239 (2003).
22. Gupta, A. S., van der Meer, M. A., Touretzky, D. S. & Redish, A. D. Segmentation of spatial experience by hippocampal theta sequences. *Nature Neurosci.* **15**, 1032–1039 (2012).
23. De Almeida, L., Idiart, M., Villavicencio, A. & Lisman, J. Alternating predictive and short-term memory modes of entorhinal grid cells. *Hippocampus* **22**, 1647–1651 (2012).
24. Bieri, K. W., Bobbitt, K. N. & Colgin, L. L. Slow and fast gamma rhythms coordinate different spatial coding modes in hippocampal place cells. *Neuron* **82**, 670–681 (2014).
25. Lee, A. M. *et al.* Identification of a brainstem circuit regulating visual cortical state in parallel with locomotion. *Neuron* **83**, 455–466 (2014).
26. O'Keefe, J., Burgess, N., Donnett, J. G., Jeffery, K. J. & Maguire, E. A. Place cells, navigational accuracy, and the human hippocampus. *Phil. Trans. R. Soc. Lond. B* **353**, 1333–1340 (1998).
27. Lever, C. *et al.* in *The Neurobiology of Spatial Behaviour* (ed. Jeffery, K. J.) (Oxford Univ. Press, 2003).
28. McNaughton, B. L., Barnes, C. A. & O'Keefe, J. The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. *Exp. Brain Res.* **52**, 41–49 (1983).
29. Czurkó, A., Hirase, H., Csicsvari, J. & Buzsáki, G. Sustained activation of hippocampal pyramidal cells by 'space clamping' in a running wheel. *Eur. J. Neurosci.* **11**, 344–352 (1999).
30. Hasselmo, M. E., Giocomo, L. M. & Zilli, E. A. Grid cell firing may arise from interference of theta frequency membrane potential oscillations in single neurons. *Hippocampus* **17**, 1252–1271 (2007).
31. Blair, H. T., Welday, A. C. & Zhang, K. Scale-invariant memory representations emerge from moire interference between grid fields that produce theta oscillations: a computational model. *J. Neurosci.* **27**, 3211–3229 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A.M. Amundsgård, K. Haugen, K. Jenssen, E. Kråkvik, and H. Waade for technical assistance, R. Báldi for help with data collection in two initial experiments, J. Couey for inspiring the bottomless car, and A. Treves for discussions. The work was supported by two Advanced Investigator Grants from the European Research Council ('CIRCUIT', Grant Agreement No. 232608; 'GRIDCODE', Grant Agreement No. 338865), the European Commission's FP7 FET Proactive programme on Neuro-Bio-Inspired Systems (Grant Agreement 600725), an FP7 collaborative project ('SPACEBRAIN', Grant Agreement No. 200873), the Kavli Foundation, the Louis-Jeantet Prize for Medicine, the Centre of Excellence scheme of the Research Council of Norway (Centre for the Biology of Memory and Centre for Neural Computation), and a PICT 2012-0548 Grant to E.K. from the Ministry of Science of Argentina.

Author Contributions E.K., M.-B.M. and E.I.M. designed experiments and analyses; E.K. and J.E.C. performed the experiments; E.K. performed the analyses; E.K. and E.I.M. wrote the paper with input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.K. (ekropff@leloir.org.ar) or E.I.M. (edvard.moser@ntnu.no).

METHODS

Subjects. Twenty-six male Long Evans rats (aged 3–6 months; 350–500 g at implantation and testing) were housed individually in transparent Plexiglass cages ($54 \times 44 \times 35$ cm). Eight of these rats were taken from a previous study of grid-cell activity during hippocampal remapping^{3,13} and one was from a study of grid cells in darkness¹. Speed cells were not reported in those studies. All rats were maintained on a 12-h light/12-h dark schedule and tested in the dark phase. After surgery, the rats were placed on a food deprivation schedule that initially kept them at ~90% of their free-feeding body weight but was progressively loosened depending on task performance.

The experiments were performed in accordance with the Norwegian Animal Welfare Act and the European Convention for the Protection of Vertebrate Animals used for Experimental and Other Scientific Purposes. The study contained no randomization to experimental treatments and no blinding. Sample size (number of animals) was initially set to five for pilot studies (linear and two-speed step protocol in the bottomless car). The number was then progressively increased up to 17 for the four-speed protocol and open-field trials. No statistical methods were used to predetermine sample size.

Electrode implantation and surgery. Tetrodes were constructed from four twisted 17- μ m polyimide-coated platinum-iridium (90–10%) wires (California Fine Wire) and mounted in groups of four into microdrives with a single turning screw and no separation between tetrodes. The electrode tips were plated with platinum to reduce electrode impedances to between 150–300 k Ω at 1 kHz.

Anaesthesia was induced by placing the animal in a closed glass box filled with isoflurane vapour. Following this, the animals were rapidly moved into the stereotaxic frame, which had a mask connected to an isoflurane pump. Air flow was kept at 1 l per minute with 0.5–3.5% isoflurane as determined by physiological monitoring. Local anaesthetic (Xylocain) was applied on the skin before making the incision. Holes were drilled on the dorsal skull anterior to the transverse sinus to reach the entorhinal cortex, and posterior to bregma to reach the hippocampus. Rats were implanted with two microdrives aiming at entorhinal cortex alone bilaterally and a third microdrive aimed at the right hippocampus. The coordinates for entorhinal implants were: 4.5–4.8 mm medio-lateral relative to lambda, 0.2–0.7 mm anterior to the border of the sinus depending on the target layer, and 1.5–1.8 mm dorso-ventral relative to the surface of the brain. The inclination of the entorhinal tetrodes was 8°, with the tips pointing in the anterior direction. Out of 13 drives that were used to record from MEC layer II, three tracks were observed to reach the very dorsal tip of the layer, at the transition to parasubiculum. After corroborating that the data from these drives was equivalent to the rest of MEC layer II in every analysed aspect (cell type proportions, theta modulation, prospectiveness of speed and grid cells), we pooled the data from all 13 drives together. The coordinates for hippocampal implants were: 2.7 mm medio-lateral, –3.3 mm antero-posterior relative to bregma, and 1.5 mm dorso-ventral relative to the brain surface. These tetrodes were implanted vertically. Jeweller's screws and dental cement were used to secure the drive to the skull. Two screws per microdrive were additionally connected to the system ground. Tetrodes in the MEC were implanted by a similar approach in the remapping and darkness study^{1,3,13}.

Data collection. For data collection, the rat was connected to the recording equipment (Axona Ltd) via a.c.-coupled unity-gain operational amplifiers close to its head, using a counterbalanced cable that allowed the animal to move freely within the available space. Tetrodes were lowered in steps of 50 μ m every day in search of new cells. All data from the same day were pooled together for cell classification, so that each cell recorded at a given depth was counted only once. In separate analyses, cells were not included if a cell had been recorded on the same tetrode at a distance of less than 200 μ m. Cell counts with these separate analyses were similar to those performed on the total data set (Extended Data Fig. 6f). Recorded signals were amplified 10,000–25,000 times and band-pass filtered between 0.8 and 6.7 kHz. Triggered spikes were stored to disk at 48 kHz (50 samples per waveform, 8 bits per sample) with a 32 bit time stamp (clock rate at 96 kHz). EEG was recorded single-ended from one electrode per drive. The EEG was amplified 5000–10,000 times, lowpass-filtered at 500 Hz, sampled at 4,800 Hz, and stored with the unit data. A tracker system (Axona Ltd) was used to record the position of a pair of LEDs attached to the head stage at a rate of 50 samples per second, allowing to track for position and head direction. The x and y components of the velocity and acceleration vectors were computed from the tracked positions using a Kalman filter and smoother (RTS).

Head oscillations, amplified by the distance of the LEDs from the head of the animal, could generate a spurious correlation between tracked acceleration and position. Anticipated spatial firing (Fig. 4d–f) could reflect such spurious correlations. In the following, we present several factors suggesting that the correlation is instead generated by genuine prospective coding. (i) The prospective nature of speed and spatial cells is not present in the hippocampus, but all hippocampal data were recorded simultaneously with data from the MEC, sharing a common

tracking signal. Moreover, in grid cells the spatial shifts were layer specific, and in speed cells prospective firing was only present in theta-modulated cells (Extended Data Fig. 9a). (ii) Under extreme conditions, firing fields shifted ~10 cm. In the worst scenario, with head oscillations of 90°, this corresponds to a distance between LEDs and head of more than 7 cm. However, LEDs were placed at most 2–3 cm away from each other, given the small size of Axona microdrives. A more realistic oscillation of 30° would require a distance of 19 cm to generate a similar field shift. (iii) The acceleration-related grid field shift was asymmetric, showing anticipation only during epochs with positive acceleration, with no effect during negative acceleration (Extended Data Fig. 10b–e). (iv) The shift was also strongly modulated by theta phase, while acceleration and theta phase were not correlated (Extended Data Fig. 10h, bottom). (v) A qualitatively similar anticipatory shift was observed in the absence of acceleration (Fig. 4b).

Spike sorting and cell classification. Spike sorting was performed offline using graphical cluster-cutting software (tint; Axona Ltd). Clustering was performed manually in two-dimensional projections of the multidimensional parameter space (consisting of waveform amplitudes), using autocorrelation and cross-correlation functions as additional separation tools and separation criteria. In general, the stability of the tetrodes allowed for all sessions in a day to be merged for clustering purposes.

Bottomless car and open field. Every day the rats were first trained in an open field (1 m \times 1 m \times 50 cm box) and then in a bottomless car on a 4 m long linear track, with possible repetition of both types of trials. There was a total of 2,010 recording sessions (total for all 26 rats). In the open field, the animal was trained to collect chocolate crumbs thrown randomly into the box, one at a time, in trials that lasted at least 20 min and as long as the animal would exhibit active foraging behaviour. Bottomless car trials varied depending on the protocol. With very few exceptions, a given rat was always trained with the same protocol. In general terms, sessions consisted of 10 to 25 runs on the linear track lasting at most 25 min. Naive rats generally explored the possibility of jumping over the limits of the car. This behaviour was discouraged by placing the animals back in the correct position inside the car. Escape attempts typically stopped after one or two runs, when rats discovered that a chocolate crumb was placed at each end of the track to motivate running. On rare occasions, for training purposes, additional ground chocolate was distributed randomly along the track. This made the rat focus on the track and prevented it from taking alternative strategies such as jumping over the car. Between runs the rat rested on the end of the track for a random interval between 10 and 20 s. A 6-s beep of increasing pitch indicated the beginning of the next run. Between trials, the rat rested on a towel in a large flower pot on a pedestal.

The bottomless car had a minimalistic design to prevent the rat from using it as a sitting platform or spatial reference frame. It was 28 cm long and 17 cm wide, with two ball-bearing wheels at each end. The car was supported by Plexiglas rails running slightly below the track along the sides (Fig. 1a). These lateral rails could barely be seen by the rat, giving support to the car without the need of lateral walls. At each end of the car was a wide mesh fence measuring 17 cm \times 16 cm to prevent the rat from moving ahead or behind the car while not obscuring their vision or sensation of velocity. A 25 W battery-powered motor (Japan Servo) under the track was attached to two sets of guide lines, each one pulling from one end of the car. A motorcycle battery was used as an isolated power source to avoid 50 Hz a.c. noise. Braided fishing line (>20 lb) was used as the guidance line. While the car was constructed in a minimalistic fashion, curtains were placed at both sides of the track and filled with a variety of salient visual cues, so as to make the laboratory the most salient spatial reference frame. Different scripts within the DacqUSB acquisition software (Axona Ltd) were used to control the motor that moved the bottomless car. The digital output of the recording system was transformed into analogue by means of a custom-built digital-to-analogue converter and fed as a control signal to the motor. To park the car consistently in the same position at the beginning of each run, two mechanical sensors were placed at the extremes of the track, and their output was fed to the digital input of the recording system.

Linear protocol. In order to secure that similar track segments were covered across a wide range of speeds, a linear relationship between speed and position was established by setting the car speed to vary exponentially with time.

Two-speed step protocol. The track was divided into two equal halves and different speeds were chosen for each half. The transition between them was sudden and occurred always in the same place.

Four-speed step protocol. This protocol was designed to obtain multiple transitions between four different speed groups: 7, 14, 21 and 28 cm s^{−1}. Every run was divided into six segments (S1–S6), three of them corresponding to the outbound run and the other three to the inbound run. S1: the speed was set at 7 cm s^{−1} for a fixed amount of time so as to cover roughly the first third of the track. S2: the speed was chosen randomly between the four options. The point for the next transition was also chosen randomly and varied from run to run within a range of ~75 cm. S3: the speed was chosen randomly again and kept until the end of the

track was reached. S4: as in S2, the speed and the transition point towards S5 were chosen randomly. S5: the speed was chosen randomly and kept until a fixed position (the same as for the transition between S1 and S2). S6: the speed was set to 28 cm s^{-1} until the end of the track. The protocol has elements of complexity that exceed the aims of this paper. For analyses of behaviour at different speeds, only segments S2 to S5, where space and speed were randomly related, were taken into account. Periods of 1 s around each transition were excluded.

The linear protocol and the two-speed step protocol were used only for visualization of speed–rate relationships, considering that space and speed were correlated in these two protocols. Further analyses were performed with the four-speed protocol, in which the two variables could be disentangled.

Remapping experiment. Recordings were obtained from eight rats while they foraged freely in enclosures in rooms A and B, following the order ABA'. In each recording session, the protocol was similar to the open field protocol described above. Enclosures could be either 1-m-wide square boxes or circular boxes 90 cm or 180 cm in diameter^{3,13}. Each trial in one enclosure lasted 10 min.

Darkness experiment. Recordings were obtained from a rat foraging freely in a circular box 1 m in diameter with the lights of the room turned either on or off, following the order ON-OFF-ON' (10 min each). Switching off the lights resulted in complete darkness.

Rate maps and speed tuning. Rate maps that showed firing rate as a function of location, head direction or speed were constructed with similar procedures. Histograms of spike count on one hand and time spent in the location on the other were built for each cell, using equally spaced bins (bin size: 2.5 cm for spatial maps, 6° for head-direction maps and 2 cm s^{-1} for speed maps). Each bin of the rate map was obtained as the ratio between the spike count and the time spent, smoothed by a Gaussian filter (standard deviation: 4 cm for spatial maps, 6° for head-direction maps and 3 cm s^{-1} for speed maps). In speed maps, where coverage is very inhomogeneous, only bins accounting for at least 0.5% of the data were included as valid. In composite rate maps for speed versus head direction, this threshold was divided by the number of head-direction bins. Instantaneous firing rate was obtained by dividing the whole session into 20-ms bins, coinciding with the frames of the tracking camera. A temporal histogram of spiking was then obtained, smoothed with a 250-ms-wide Gaussian filter. Spatial and head-directional information measures³² were based on these maps. The variability of a cell's speed map or tuning curve when comparing two sessions A and B (Fig. 3i) was calculated as the average across bins of the absolute normalized change in firing rate, $\frac{|(A-B)|}{|(A+B)|}$.

The speed score for each cell was defined as the Pearson product-moment correlation between the cell's instantaneous firing rate and the rat's instantaneous running speed, on a scale from -1 to 1 .

Shuffling. Chance-level statistics was constructed for a given variable W through a shuffling procedure. At each repetition, the entire sequence of spikes fired by the cell was time-shifted along the animal's path by a random interval between 30 s and the total trial length minus 30 s, with the end of the trial wrapped to the beginning. The shuffled instance of the variable W was then calculated using the shifted spikes, and the collection of 100 repetitions for each cell composed the chance-level statistics. For cell-type classifications, all shuffled data of the corresponding score was pooled together and the 99th percentile of the distribution was used as a classification criterion.

In order to distinguish speed-correlated effects from changes in behavioural state (foraging versus sitting still), we dismissed all data produced at a running speed lower than 2 cm s^{-1} in the calculation of the observed and shuffled speed scores.

Normalization of speed cell activity. Because of the variability in baseline and slope, a simple or normalized average of speed cell activity would not properly capture the population behaviour. To obtain a better normalization method, we applied to any firing rate measure f of a speed cell expressed in Hz the linear transformation

$$f_n = \frac{(f-A)}{B \times 50}$$

where A (Hz) and B (cm^{-1}) are the y intercept and slope of the cell's speed tuning. The 50 value is given in cm s^{-1} . This linear transformation aims to achieve for every cell a normalized dimensionless activity of 0 when the rat is still and 1 when it runs at 50 cm s^{-1} , allowing for proper population averaging.

Unbiased analysis of cells modulated by running speed. The modulation depth of a cell was defined as the difference between the maximum and minimum firing rates in its speed-tuning curve. A cell was classified as modulated by speed if its modulation depth was significantly higher than the 99th percentile of a distribution of modulation depths obtained from shuffling 1,000 times the cell's spike time stamps. It is worth noting that the nature of the modulation depth does not allow for the mixture of information coming from different cells, so that every

individual cell had its own threshold. This selection method has no bias towards linear coding of speed, but for all types of data a majority of significantly modulated cells exhibited a positive, linear code, as measured by the linearity index (regression of the tuning curve) (Extended Data Fig. 3b).

Measures used for cell type classification. *Gridness score*^{13,15,16}. The gridness score for each cell was determined from a series of expanding circular samples of the autocorrelogram, each centred on the central peak but with the central peak excluded. The radius of the central peak was defined as either the first local minimum in a curve showing correlation as a function of average distance from the centre, or as the first incidence where the correlation was under 0.2, whichever occurred first. The radius of the successive circular samples was increased in steps of 1 bin (2.5 cm) from a minimum of 10 cm more than the radius of the central peak, to a maximum of 90 cm. For each sample, we calculated the Pearson correlation of the ring with its rotation in α degrees first for angles of 60° and 120° and then for angles of 30° , 90° and 150° . We then defined the minimum difference between any of the elements in the first group (60° and 120°) and any of the elements in the second (30° , 90° and 150°). The cell's gridness score was defined as the highest minimum difference between group-1 and group-2 rotations in the entire set of successive circular samples.

*Mean vector length (head-direction score)*³³. Given the head-direction tuning map of a cell, if the bin i with orientation θ_i expressed in radians is associated with a firing rate λ_i , the mean vector length was computed as

$$\left| \frac{\sum \lambda_i e^{i\theta_i}}{\sum \lambda_i} \right|$$

where the sums were performed over all N directional bins and the modulus of the resulting complex number was obtained.

*Information per spike*³². Given a spatial or head-direction map with mean firing rate λ and a value λ_i for each of its N bins, information rate was computed as

$$\sum_{i=1}^N p_i \frac{\lambda_i}{\lambda} \log_2 \left(\frac{\lambda_i}{\lambda} \right)$$

where p_i is the occupancy probability of bin i .

Border score^{17,18}. The border score was computed as the difference between the maximal length of a wall touching on any single firing field of the cell and the average distance of the field from the nearest wall, divided by the sum of those values. The range of border scores was thus -1 to 1 . Firing fields were defined as collections of neighbouring pixels with firing rates higher than 20% of the cell's peak firing rate and a size of at least 200 cm^2 .

*Theta index*³⁴. For a given cell, the normalized temporal autocorrelogram was obtained using bins of 5 ms. The theta index was defined as the difference between the trough (50–70 ms) and the peak (100–140 ms).

Estimation of the significance of overlaps between cell populations. The observed population overlaps were compared with the ones that would result from an independent random assignment of categories. The probability of a neuron to be randomly assigned to category A was set as $p_A = N_A/N$, where N is the total number of neurons and N_A the total number of neurons belonging to A. In this way, in a population of N neurons the expectation value of the size of the subpopulation randomly assigned to category A is $p_A \times N = N_A$, identical to the observed group size. Since the assignments are random and independent, the probability of a neuron to be assigned simultaneously to categories A and B is $p_{AB} = p_A \times p_B$, and the expectation value for the overlap between both populations is $p_{AB} \times N$. When speed was one of the categories, the observed overlap N_{AB} was consistently found to be lower than $p_{AB} \times N$ (Extended Data Fig. 6c). To estimate the significance of this difference, N_{AB} was compared with the full probability distribution. In a Bernoulli process, the probability of succeeding k times when tossing N times a coin, each time with probability of success p_{AB} , is given by the binomial distribution

$$p(k) = \binom{N}{k} p_{AB}^k (1-p_{AB})^{N-k}$$

and the left tail P value associated to N_{AB} is

$$P = \sum_{k=0}^{N_{AB}} p(k).$$

Decoding of running speed from speed cell activity. A simple linear decoder was implemented³⁵. A linear relationship between firing rate and speed averaged over 1-s bins is expressed as

$$S_{tr} = R_{tr} f$$

where S_{tr} is a column vector with the speed bins used for training, R_{tr} is a matrix containing, as columns, the corresponding bins of firing rate for each neuron and

an additional column of 1s to account for y intercepts, and f is the linear filter, also a column vector, with length equal to the number of neurons plus 1. Training the filter is equivalent to inverting this equation,

$$f = (S_{tr}^T S_{tr})^{-1} S_{tr}^T R_{tr}$$

where T and -1 indicate the transpose and the inverse of a matrix, respectively. Once f is obtained, the reconstructed speed (S_{rec}) for a different set of firing rates R_{test} of the same neurons is obtained as

$$S_{rec} = R_{test} f$$

and the reconstruction quality is defined as the Pearson correlation between S_{rec} and the actual speed S_{test} .

Fields on the linear track. Grid and place fields on the linear track were individualized from one-dimensional spatial maps, treating outbound and inbound runs separately. Fields were identified as isolated local maxima in the rate map above 2 Hz, decaying at both sides to either half of their amplitude or below 2 Hz before a new local maximum appeared. A Gaussian fit around the peak of the field was used to estimate amplitude, centre and standard deviation of the field. The z score for any position on the track was defined as its distance to the closest field centre divided by the standard deviation of the field. The sign of the z score was adjusted for inbound and outbound runs such that the running direction always went from negative to positive values.

For a quantification of spatial shifts in real space (Fig. 4b) the position relative to the field centre rather than the z score was used. The two measures are different only in the normalization by field standard deviation. Gaussian fits were used to estimate the field centres subject to different running speeds.

Quantifying the temporal anticipation of the grid field. We define two different sets of kinetic variables. The position, speed and acceleration of the rat are represented by $x(t)$, $v(t)$ and $a(t)$, respectively. The same quantities calculated by a prospective path integrator (which we assume to be free of errors) are represented instead by $\tilde{x}(t)$, $\tilde{v}(t)$ and $\tilde{a}(t)$, respectively. For simplicity and without loss of generality, we assume all these quantities to be zero at time $t = 0$.

If prospective speed cells anticipate the running speed by τ , we can write

$$\tilde{v}(t) = v(t + \tau) = v(t) + \tau a(t) + \frac{\tau^2}{2} \frac{\partial a(t)}{\partial t} + \dots \quad (1)$$

where we have used the Taylor series expansion of $v(t + \tau)$ around t . The position of the animal at any time t can be expressed as

$$x(t) = \int_0^t v(u) du.$$

However, a prospective path integrator that used $\tilde{v}(t)$ as a speed signal would calculate the position at time t as

$$\tilde{x}(t) = \int_0^t \tilde{v}(u) du = x(t) + \tau v(t) + \frac{\tau^2}{2} a(t) + \dots$$

where we have used (1).

In the four-speed experiment, we can choose to work with segments of constant running speed, where the acceleration and all other derivatives of the speed are close to zero. Thus, a grid field will suffer a spatial anticipation following

$$\tilde{x}(t) \approx x(t) + \tau v(t) \quad (2)$$

Intuitively, if the anticipation of the grid field is of a temporal nature, it will be seen in space as proportional to the running speed, with τ as the coefficient of proportionality.

Acceleration-related field shift on the linear track. On the linear track, outbound and inbound runs were treated as different sessions. In order to pool together in the analysis fields with different width, the z score rather than the position on the track was used as the spatial variable, defining the running direction always from negative to positive values. For every identified field, positive and negative acceleration maps were constructed by filtering only segments of the trajectory with the corresponding acceleration sign and where absolute acceleration passed a pre-set threshold, for example, 50 cm s^{-2} . The variable Δ was defined as the spatial shift of the positive map that maximized its correlation with the negative map. A field was considered for further analysis only when the correlation between positive and negative maps at its maximizing shift Δ was above 0.9.

This measure did not allow for the dissection of the prospective and retrospective components of the shift, assuming alternating modes²³. To make this distinction, the average firing field was used as a reference. We used only experiments where this reference could be assumed to be nearly unbiased, that

is, the four-speed protocol, where most of the time was spent at the lowest speed (7 cm s^{-1}) and acceleration was close to zero. In this case, Δ was defined as the spatial shift of the positive or negative map that maximized its correlation with the reference field. Positive (negative) values of Δ along the running direction characterized prospective (retrospective) coding.

Open-field acceleration. While in 1D the sign of acceleration is always well defined, in 2D this happens only when the acceleration vector points approximately in the same or in the opposite direction of the velocity vector. These vectors, obtained from their x and y components of the position, were decomposed into magnitude (a and v) and direction (a_d and v_d). All open-field analyses considering the sign of acceleration include only segments of the trajectory where the absolute value of $\cos(a_d - v_d)$ is greater than 0.8. This excludes deviations greater than $\sim 37^\circ$. The effective acceleration was thus defined through its magnitude a and the sign given by the sign of $\cos(a_d - v_d)$.

Field shift in the open field. Since in open-field experiments spatial fields are never traversed twice in the same way, a map-based method was developed to estimate field shift caused by acceleration. The running direction of the rat was used to divide the session where a place cell or grid cell was recorded into four groups: north, east, west and south. Different groups were treated as if they were independent sessions. For every acceleration threshold with an absolute value of a_t , two spatial maps were constructed, selecting only time stamps where acceleration was well defined, greater than a_t in absolute terms and either positive or negative. The first of these maps was then displaced to both sides in the running direction (north, east, west or south) in order to determine the displacement Δ that maximized the correlation between both maps. Only maps with a maximum firing rate above 10 Hz and with a maximum correlation between maps above 0.5 were included in the analysis. All cells in all four running directions were pooled together for the population analysis.

Theta rhythm. A band-pass filter with cut-off frequencies of 6 Hz and 12 Hz was applied to the raw EEG data in order to obtain the theta component of the local field potential. A Hilbert transform was used to decompose the resulting oscillation into amplitude and phase. The phase was then unwrapped into a mostly monotonically increasing signal by adding 2π at every phase reset. The phase at which spike and position time stamps occurred was obtained from the unwrapped phase by interpolation followed by a modulo 2π operation. These values were used to construct histograms of phase precession in the space versus theta phase domain. The theta index³⁴ was used to assess the theta modulation of individual neurons.

Clustering of theta-phase related behaviour. The theta cycle was binned (16 bins) and the average normalized firing rate of each speed cell for each bin was obtained. This data was used as an input into a k -means clustering algorithm (MATLAB) with the number of clusters k varying between four and ten. In every case we used the best result out of ten replicates, defined as the one with the lowest within-cluster sum of point-to-centroid distance, which ensured a stable solution. Four qualitatively different behaviours were consistently found. For all values of k greater than four, the 'ramping' cluster split into sub-clusters of ramping activity with different grades of steepness. After, merging these sub-clusters into one, solutions with different values of k were very similar to each other. The results in Extended Data Fig. 9 use $k = 7$.

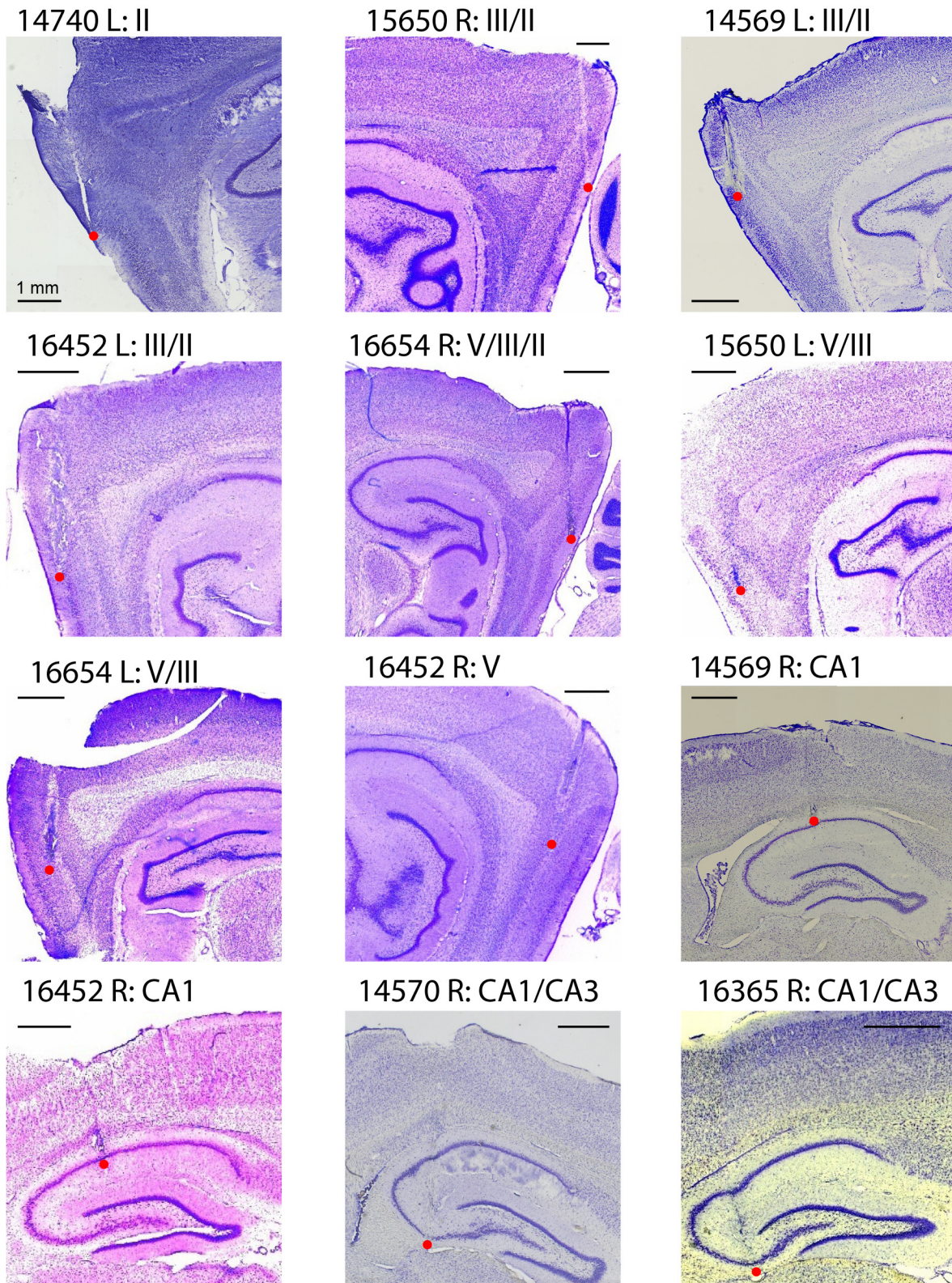
Histology. Electrodes were not moved after the final recording session. Anaesthesia was induced by placing the animal in a closed glass box filled with isoflurane vapour. The rats then received an overdose of Equithesin and were perfused intracardially with saline and 4% formaldehyde. The brains were extracted and stored in formaldehyde, and frozen sagittal sections ($30 \mu\text{m}$) were cut. All sections were mounted on glass slides and stained with cresyl violet. With the use of a light microscope, equipped with a digital camera, the positions of the recording electrodes were registered in relation to relevant borders between sub-fields. Final positions of the recording electrodes were indicated on photomicrographs obtained in AxioVision. The exact position of the electrodes at recording was extrapolated using the read-out of the tetrode turning protocol.

Statistical tests. Statistical tests were two-sided and non-parametric.

Code availability. Code for obtaining smooth speed and acceleration measures can be provided by the authors.

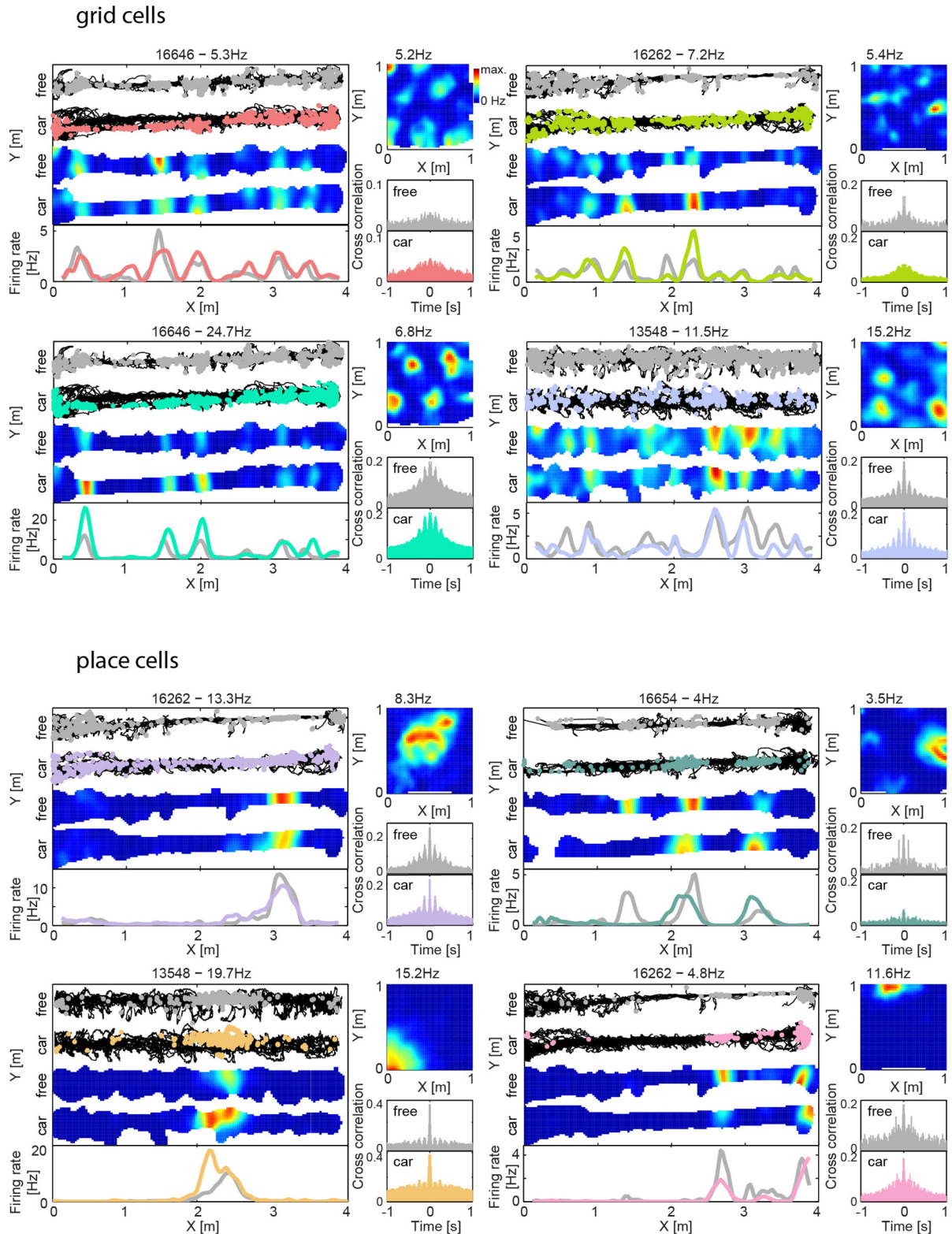
32. Skaggs, W. E., McNaughton, B. L., Wilson, M. A. & Barnes, C. A. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* **6**, 149–172 (1996).
33. Boccara, C. N. *et al.* Grid cells in pre- and parasubiculum. *Nature Neurosci.* **13**, 987–994 (2010).
34. Cacucci, F., Lever, C., Wills, T. J., Burgess, N. & O'Keefe, J. Theta-modulated place-by-direction cells in the hippocampal formation in the rat. *J. Neurosci.* **24**, 8265–8277 (2004).
35. Serruya, M. D., Hatsopoulos, N. G., Paninski, L., Fellows, M. R. & Donoghue, J. P. Instant neural control of a movement signal. *Nature* **416**, 141–142 (2002).
36. Terrazas, A. *et al.* Self-motion and the hippocampal spatial metric. *J. Neurosci.* **25**, 8085–8096 (2005).

37. Paninski, L., Fellows, M. R., Hatsopoulos, N. G. & Donoghue, J. P. Spatiotemporal tuning of motor cortical neurons for hand position and velocity. *J. Neurophysiol.* **91**, 515–532 (2004).
38. Chapin, J. K., Moxon, K. A., Markowitz, R. S. & Nicolelis, M. A. Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature Neurosci.* **2**, 664–670 (1999).
39. Burwell, R. D. & Amaral, D. G. Cortical afferents of the perirhinal, postrhinal, and entorhinal cortices of the rat. *J. Comp. Neurol.* **398**, 179–205 (1998).
40. O'Keefe, J. & Recce, M. L. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* **3**, 317–330 (1993).
41. Hafting, T., Fyhn, M., Bonnevie, T., Moser, M. B. & Moser, E. I. Hippocampus-independent phase precession in entorhinal grid cells. *Nature* **453**, 1248–1252 (2008).



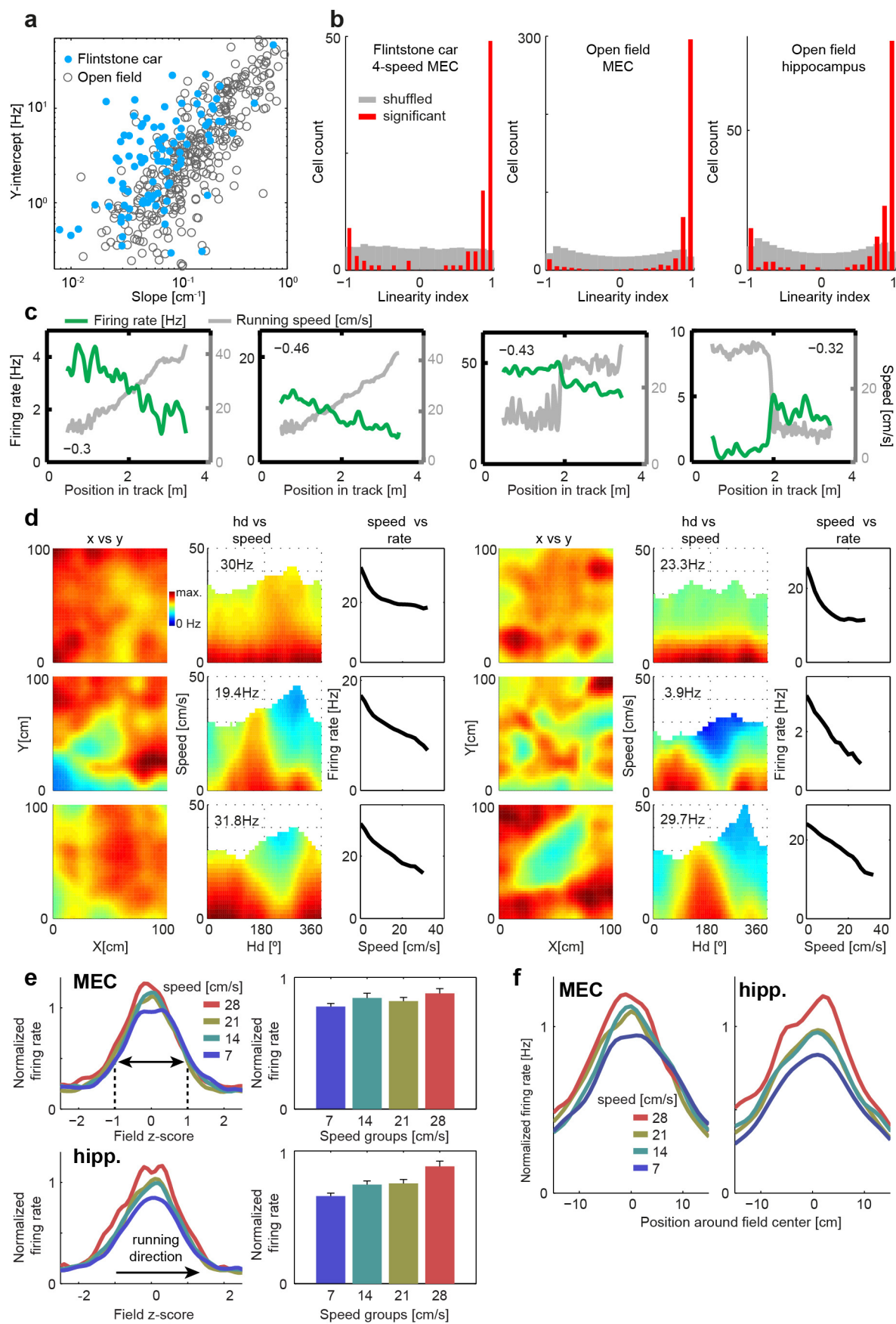
Extended Data Figure 1 | Nissl-stained sagittal brain sections showing representative recording locations in the MEC and hippocampus. Red dots indicate final location of tetrodes. Rat number, hemisphere (R, right; L, left) and

entorhinal layers or hippocampal regions where cells were recorded are indicated. Scale bars, 1 mm.



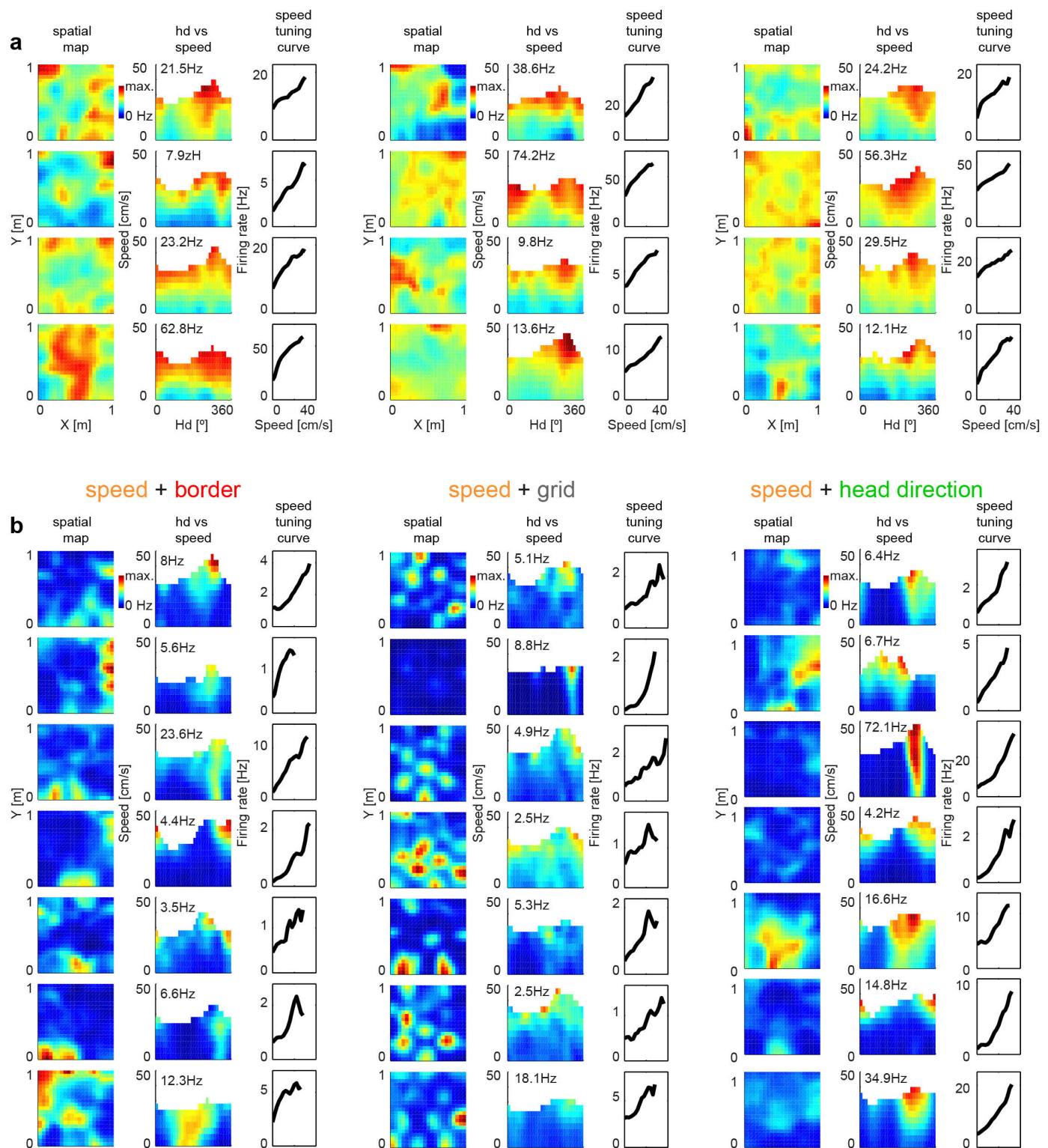
Extended Data Figure 2 | The bottomless car does not affect firing properties of grid and place cells on the linear track. As opposed to recording from a passive rat sitting on a classical car³⁶, the bottomless car task does not alter the spatial and temporal firing properties of grid cells (top, four cells) or place cells (bottom, four cells). Every cell was recorded under three conditions: experimenter-determined running in the bottomless car ('car'); free foraging on the same linear track but with the bottomless car removed ('free'); and open field. Each block of panels shows data for one cell. Left side of each panel: from top to bottom, the animal's trajectory (black curve) and spike

positions (coloured dots) for free sessions and car sessions; corresponding colour-coded rate maps, with red indicating peak rate and dark blue indicating silence; and overall firing rate across the x dimension of the track for free (grey) and car (colour) conditions. Note the similarity between spatial maps recorded in the car and the free condition. Right side of each panel: from top to bottom, colour-coded open field rate map and temporal cross-correlograms of spiking in free and car conditions. Note the similarity of the two cross-correlograms.



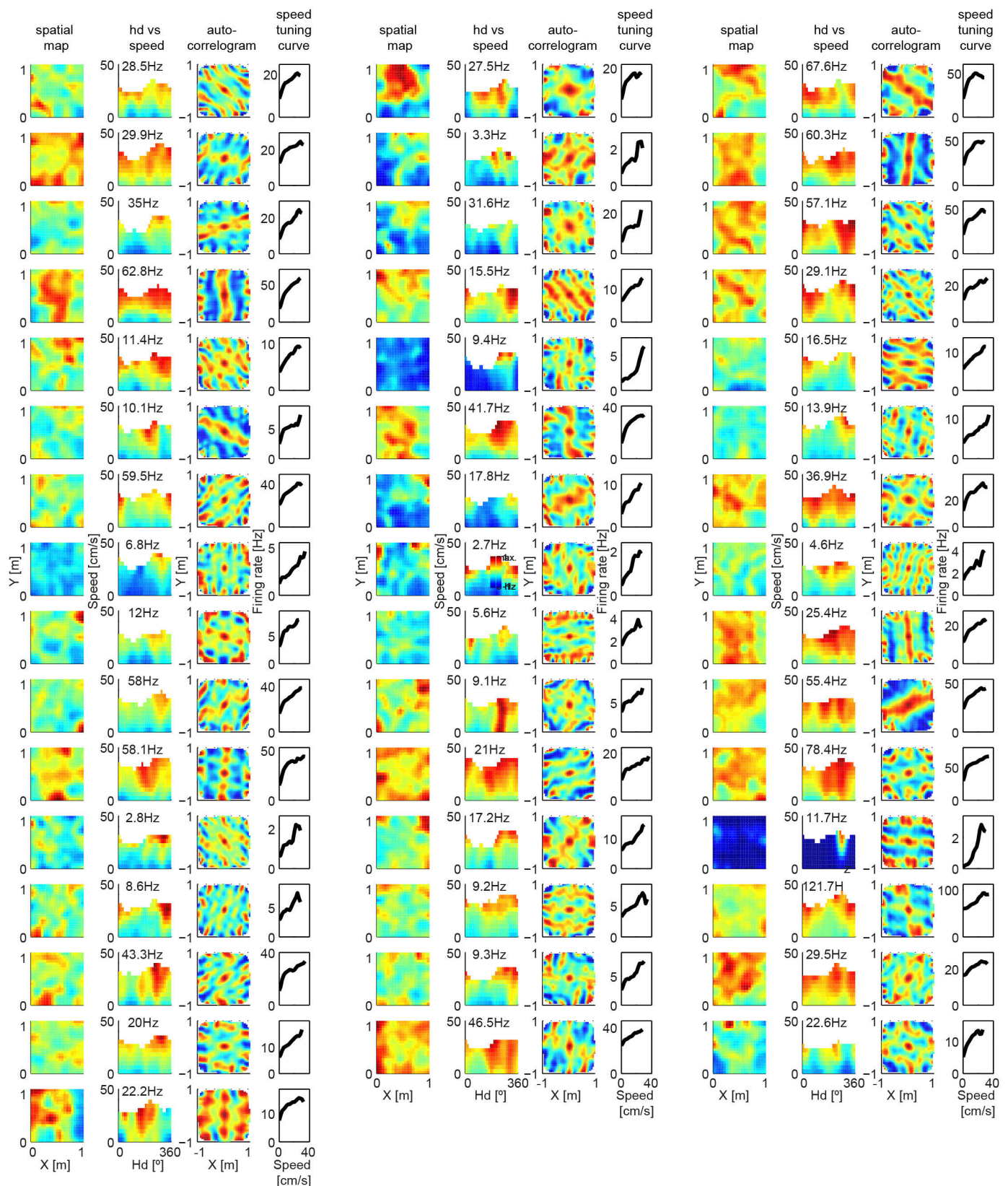
Extended Data Figure 3 | Linear relationship between speed and firing rate in speed cells but not spatially modulated cells of the MEC or the hippocampus. **a**, Scatter plot showing slope and y intercept of regression lines for each entorhinal speed cell recorded in the bottomless car (blue circles) and in the open field (grey circles). Note wide range of slopes and y intercepts. **b**, Identification of speed-modulated cells using analyses that do not assume linearity (see Methods). The linearity of these cells is represented by the regression of the tuning curves (red), which clusters mostly around 1 (speed cells) and marginally around -1 (anti-speed cells), in contrast with the distribution of linearity indexes of the shuffled population (grey, 100 shuffling steps, count normalized by the number of steps). This holds across experimental protocols and brain regions, as indicated. **c**, Spatial maps and average speed along the track of four representative anti-speed cells in the bottomless car under linear or two-speed step protocols, plotted as in Fig. 1b. **d**, Firing rate as a function of position, head direction (hd), and running speed for six representative anti-speed cells recorded in the MEC during free running in a square open field. Each row shows one cell. Left, colour-coded spatial rate maps. Scale bar to the right. Middle, firing rate as a function of head

direction (x axis) and running speed (y axis). Firing rates in left and middle diagrams share the same colour code. Right, firing rate as a function of running speed. **e**, Speed modulation of firing fields in the MEC (top) and hippocampus (bottom). Left, average normalized firing profile of fields in each of the four speed groups in the bottomless car. Right, for each field, the area under the curve 1 s.d. around the average field centre is computed to obtain mean firing rate across firing fields for each speed group (mean \pm s.e.m.). Statistical tests showed no significant effect of speed on the average normalized firing rate in the MEC (Kruskal–Wallis test, $P = 0.12$). In the hippocampus, in contrast, the same tests showed a significant trend in the modulation by speed, due exclusively to the difference between 7 and 28 cm s⁻¹ (Kruskal–Wallis and Tukey–Kramer tests, $P < 0.05$). Note that similar tests on entorhinal speed cells (Fig. 1f) showed significant differences between all groups ($P < 0.01$). **f**, Average firing fields, as in **e**, but using position relative to field centre instead of field z score as the spatial variable (running is always from negative to positive values). This allows direct measurement of firing position as a function of running speed, connecting equation (2) with Fig. 4b. Gaussian fits are used to determine firing position, defined as the field centre for each speed category.



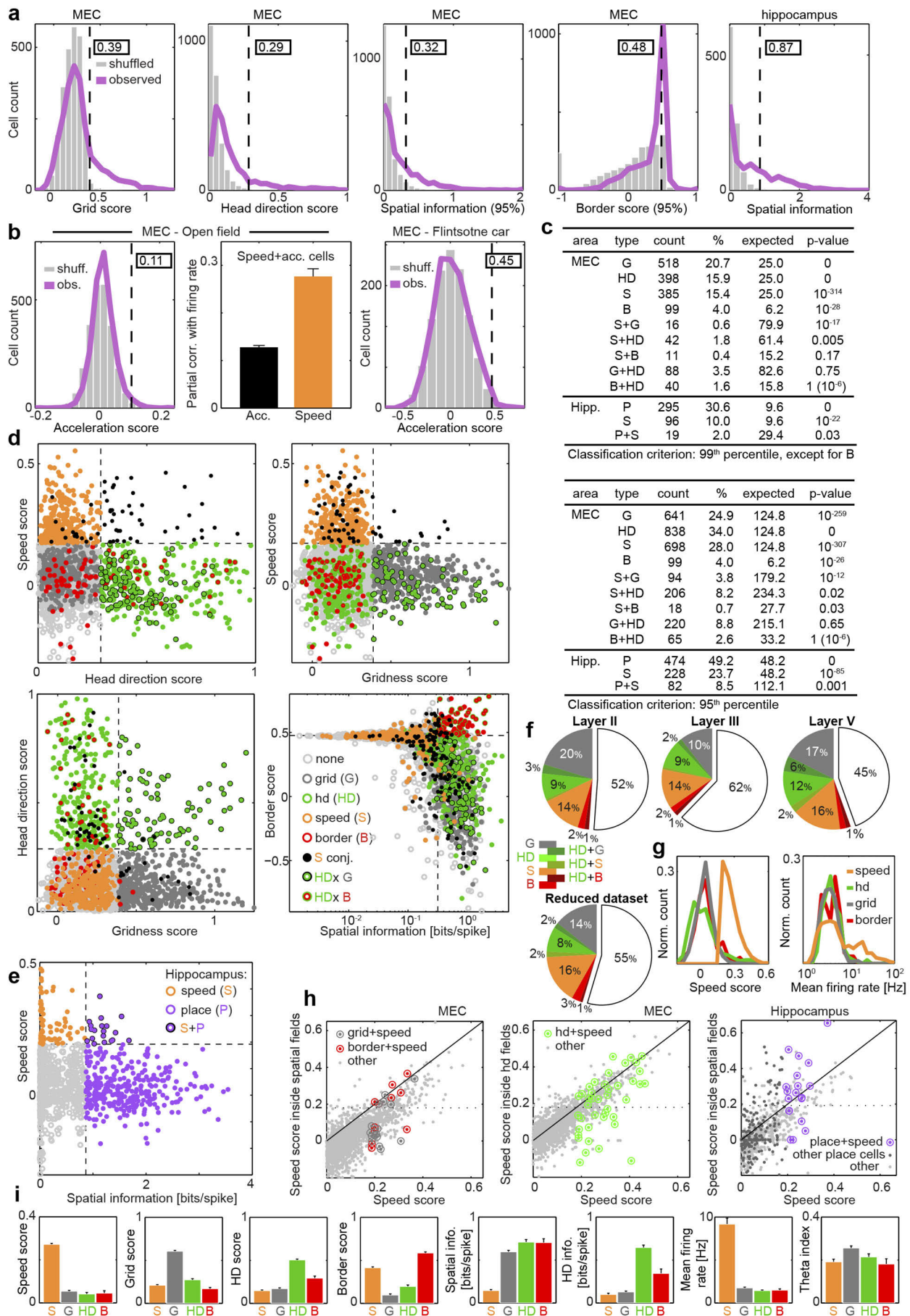
Extended Data Figure 4 | MEC speed cells are generally not modulated by space or direction. Examples of MEC speed cells. Three sets of data are shown for each cell. Left, colour-coded spatial rate maps. Scale bar to the right of the first map. Middle, firing rate as a function of head direction (x axis) and running speed (y axis). Same colour code as for the rate map. Maximum firing

rate is indicated in the upper left corner. Right, firing rate as a function of running speed. **a**, Twelve representative MEC speed cells (from a total sample of 385 speed cells in 17 animals), which in general are poorly modulated by space or direction. **b**, Examples of speed cells that passed one additional cell-type criterion (from left to right: border, grid, head direction).



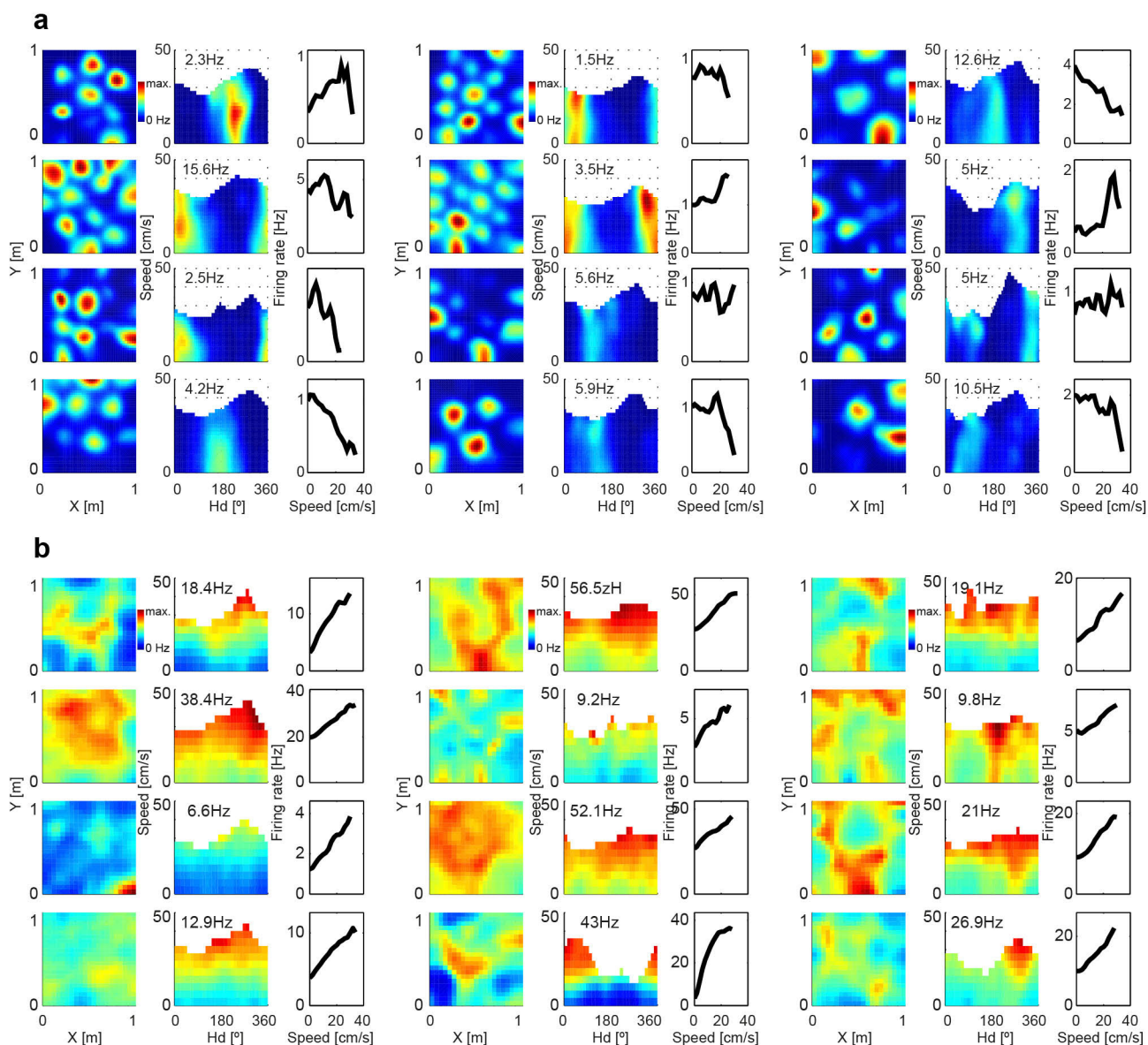
Extended Data Figure 5 | MEC speed cells in a single animal. All MEC speed cells recorded in one animal (rat 14740). For each cell we show four plots. From left to right: spatial rate maps; head-direction versus speed rate maps; spatial autocorrelograms used to calculate the gridness score; and speed tuning

curves (right). Symbols for rate map, head-direction versus speed map, and tuning curve as in Extended Data Fig. 4. The spatial autocorrelogram is colour-coded from $r = -1$ (blue) to $r = +1$ (red).



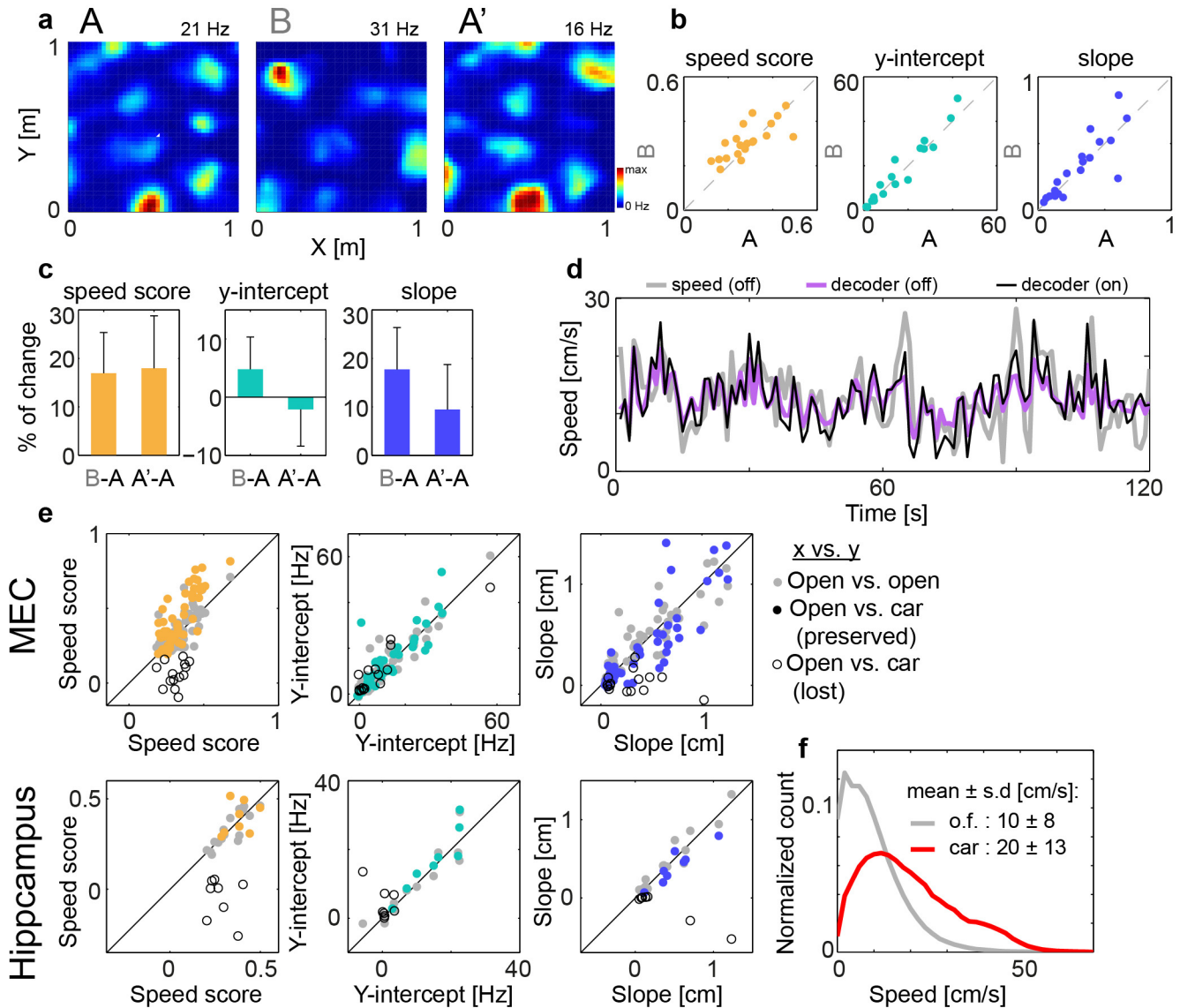
Extended Data Figure 6 | Speed cells form a separate cell class. **a**, Observed data (purple) and 100 step-shuffled distributions (grey; count normalized by 100) of different variables used to classify cell types. The dashed lines represent the 99th percentile threshold of the shuffled distribution, with the exception of the distributions of border score and spatial information used for border cell classification, where a dual 95th percentile criterion was used. Threshold values are indicated in boxes. **b**, A similar comparison with shuffled data shows no signs of 'acceleration cells' in the MEC. The acceleration score was defined as the correlation between instantaneous firing rate and acceleration. Left, cells recorded in the open field had a distribution of acceleration score (purple) very similar to that of the shuffled population (grey bars). The number of cells exceeding the 99th percentile of the shuffled distribution (0.11) were 21 more than the average chance level (observed, 46 out of 2497; expected, 25; $P = 10^{-4}$). This might be explained by the fact that out of these 46 cells, 20 were speed cells, which are as a population modulated by acceleration due to their prospective nature (Fig. 4c and equation (1)). Middle, the partial correlation between firing rate on one side and speed and acceleration on the other was computed for those speed cells with high acceleration modulation. In all cases, the partial correlation with speed was higher than the partial correlation with acceleration, with more than a twofold difference on average. Right, potential modulation by acceleration was also studied by restricting the calculation of the acceleration score to fragments of 2 s around the onsets for the highest speed change in the four-speed experiment (from 7 to 28 cm s⁻¹), where potential 'acceleration cells' should exhibit a peak in their firing rate. Cells recorded in this experiment had a distribution of acceleration scores (purple) very similar to that of the shuffled distribution (grey), and only 8 out of 997 cells had a score above the 99th percentile of the shuffled distribution (0.45; expected, 10; $P = 0.78$). **c**, Tables showing the significance of population sizes and population overlaps using classification thresholds based on the 99th (top) and the 95th (bottom) percentile of the shuffled distribution. G, grid cells; HD, head-direction cells; S, speed cells; B, border cells; P, place cells; +, conjunctive cells satisfying criteria for more than one cell class. Expected chance levels are obtained from Bernoulli distributions. For single categories, the right tail P value is indicated. For overlap between categories, the left tail P value is indicated, while in the case of the overlap between head-direction and border cells, which clearly exceeds chance levels (~40% of border cells are also head-direction cells), the right tail P value is added in parentheses. The mixture in the coding of speed and other behavioural variables was always smaller than the mixture between spatial and directional coding. For hippocampal data, the statistics include only cells that were active in the open field (not including sleep sessions). Note that all cell categories are defined by comparison with a shuffled distribution, that is, not by applying arbitrary

thresholds. This procedure does not always define populations of significant magnitude (see **b**) and exhibits consistent results for the overlap of populations at the 99th and 95th percentile level. **d**, Scatter plots showing distributions of scores and cell-type classifications. Each dot represents a cell, with the same colour code as used in Fig. 2g. x and y axes show scores used for cell-type classification (gridness score, speed score, mean vector length head-direction score, border score, or spatial information). Dashed lines represent the classification threshold for each score. **e**, Scatter plot as in **d** showing overlap between the speed-cell and the place-cell populations in the hippocampus. In this case, speed score and spatial information were used for classification. **f**, Top row, pie charts showing distribution of functional cell types and their overlaps across entorhinal layers (only proportions higher than 1%). Bottom, recording across multiple days can generate an unwanted bias in the estimation of population sizes, since a single cell could be counted many times. To avoid this bias, we reduced our original data set by discarding a cell if another cell had been recorded at a distance of less than 200 μ m on the same tetrode on an earlier day. In this reduced population of 608 cells, 18% were speed cells, confirming that the population size estimation is free of this kind of bias. **g**, Distribution for different cell categories of speed score (left) and firing rate averaged over non-silent periods (firing rate > 1 Hz; right). **h**, The speed scores of cells in the MEC (left and middle) and hippocampus (right) were plotted against the in-field speed score of the cells, calculated only with data from the bins with a firing rate above the median. This quantity is a correction for spatially and directionally modulated cells, but has no meaning for other cells. Left, out of 16 grid cells that passed the speed cell criterion, 11 (69%) had in-field speed scores clearly below threshold, while the remaining population had similar regular and in-field scores (Mann–Whitney U -test, $P = 0.31$). Similarly, out of 11 border cells, 5 (45%) had very low in-field scores, and the remaining had similar regular and in-field scores ($P = 0.82$). Middle, a similar approach was implemented using head-direction bins instead of spatial bins. Out of 42 MEC head-direction cells with high speed score, 17 (40%) had in-field scores below threshold, while the remaining population had similar regular and in-field scores ($P = 0.57$). Right, different conclusions were obtained in the analysis of hippocampal place cells. Out of 19 place cells with high speed score, 6 (32%) had low in-field scores. The remaining population had in-field scores significantly higher than the corresponding regular speed scores (Mann–Whitney U -test, $P < 0.02$). In addition, 33 other place cells with low regular speed score had in-field speed scores higher than threshold, suggesting a stronger mixture between speed and spatial coding in the hippocampus. **i**, Population distribution (mean \pm s.e.m.) of various quantities for all MEC cell types (S, speed; G, grid; HD, head direction; B, border).



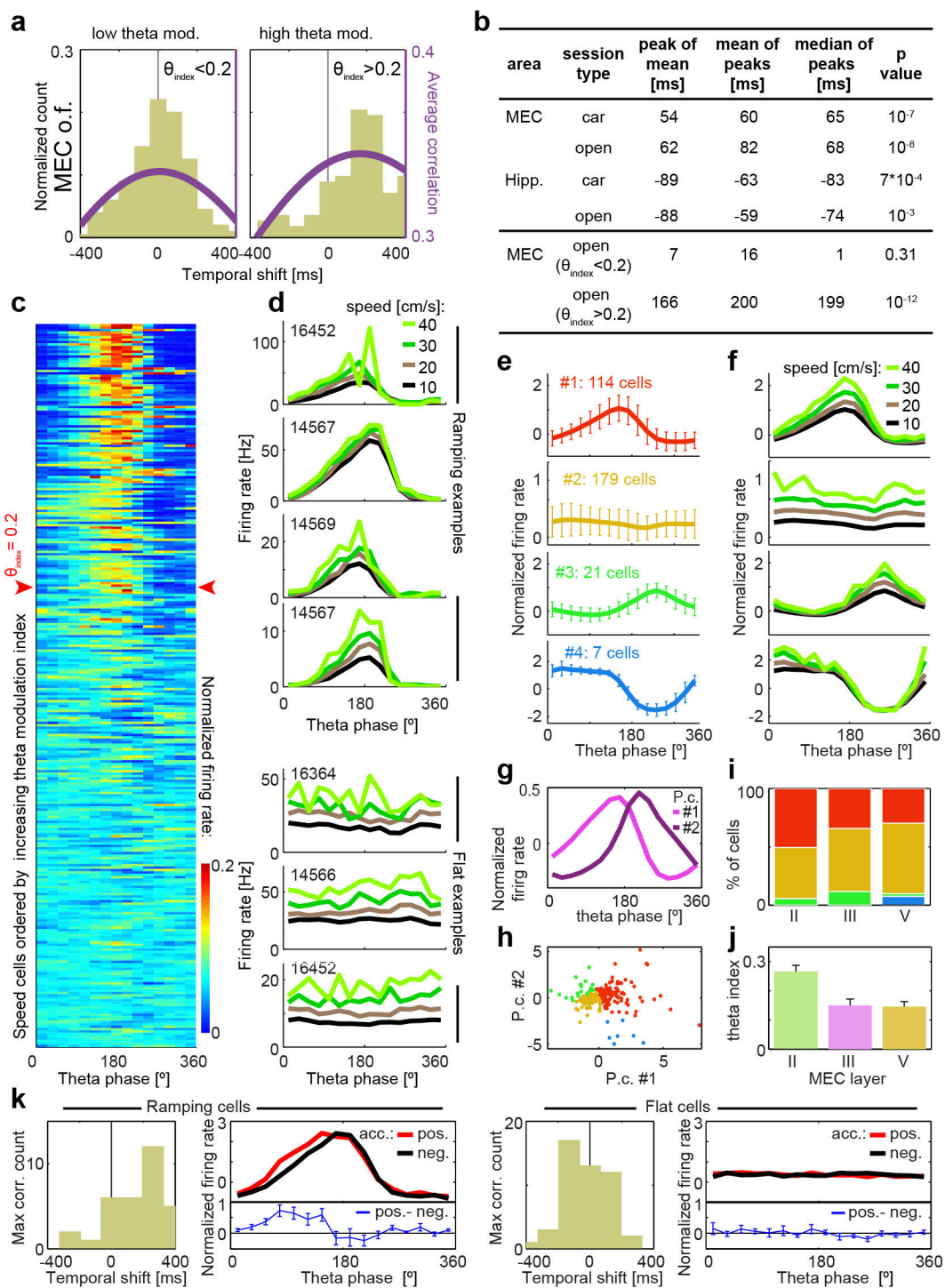
Extended Data Figure 7 | Representative examples of conjunctive grid and head-direction cells in the MEC and speed cells in the hippocampus. Three sets of data are shown for each cell. Left, colour-coded spatial rate maps. Scale bar to the right of the first map. Middle, firing rate as a function of head

direction (x axis) and running speed (y axis). Same colour code as for the rate maps. Right, firing rate as a function of running speed. **a**, MEC conjunctive cells do not exhibit strong modulation by speed. **b**, Hippocampal speed cells have characteristics that are similar to entorhinal ones.



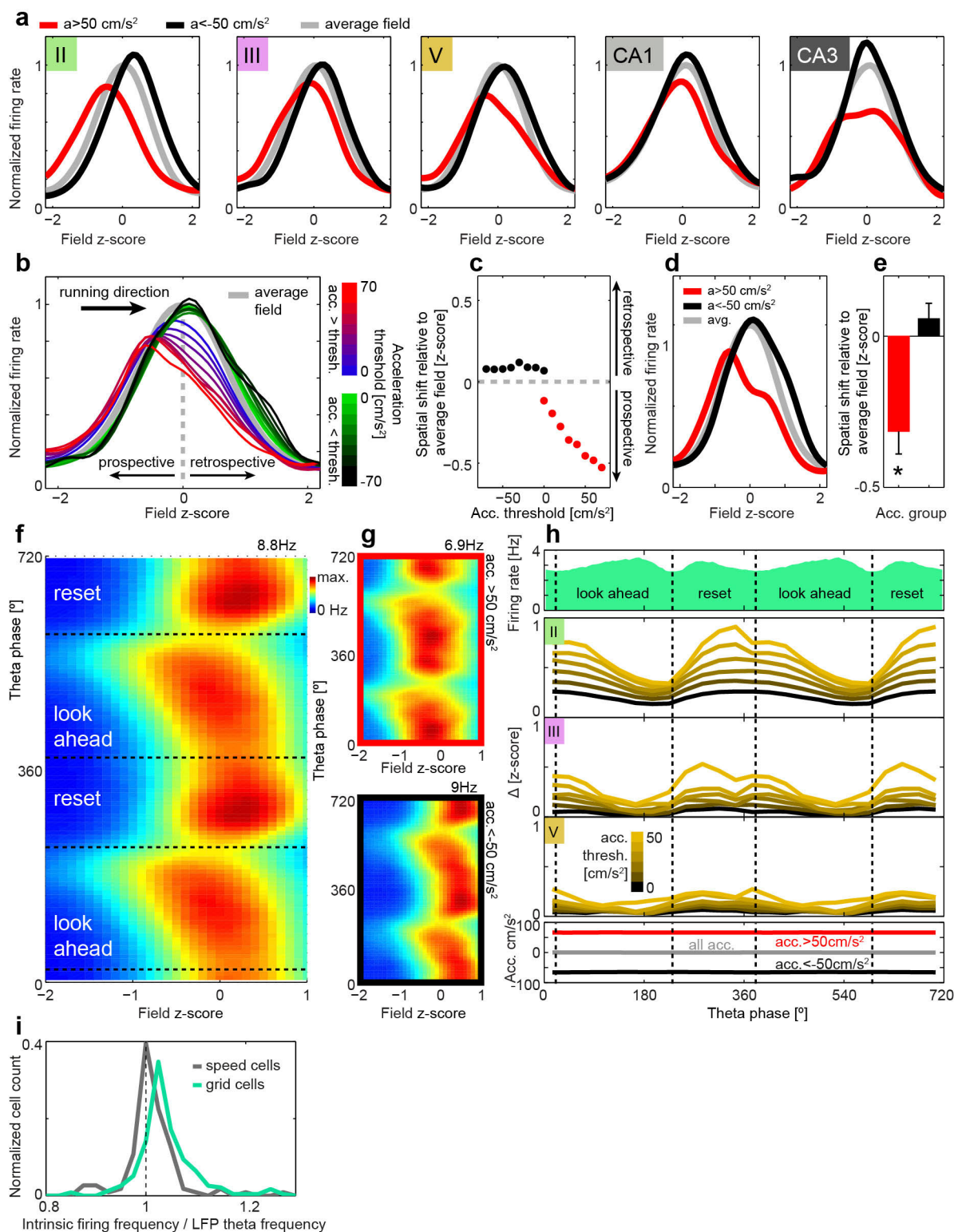
Extended Data Figure 8 | The speed code is context-invariant. **a**, Colour-coded rate maps showing realignment in a grid cell recorded in rooms A and B. The sequence of recording was ABA'. In the MEC, change of room causes change in grid phase and grid orientation; in the hippocampus, this is accompanied by global remapping³. **b**, Speed score, tuning curve y intercept and slope in room A versus room B for 20 speed cells recorded in the room-change experiment in **a** (eight rats). Each dot corresponds to one cell. Values distributed around the diagonals indicate context invariance. **c**, Percentage change for the same quantities between trials A and B and between A and A' (mean \pm s.e.m.). In each case, the difference between the two distributions was non-significant (Wilcoxon signed rank test, speed score, $P = 0.9$; y intercept, $P = 0.54$; slope, $P = 0.49$). **d**, Reconstructed speed (purple and black) compared to actual speed in darkness (grey). Speed was decoded from the activity of three speed cells (Fig. 3f, g), with decoders trained either in the lights-on condition (black) or the lights-off condition (purple). Pearson correlation between reconstructions was 0.97. Correlation between decoded speed and actual speed was 0.45 with the 'light on' decoder, and 0.48 with the 'light off' decoder. **e**, All speed cells that were recorded in the open field both before

and after trials in the bottomless car were selected from recordings in the MEC (top) and the hippocampus (bottom). Speed score (left), speed tuning curve y intercept (middle) and slope (right) were compared within sessions. Grey dots show the comparison between pre-open-field and post-open-field recordings (x axis and y axis, respectively). Coloured circles indicate the comparison between pre-open-field (x axis) and bottomless car (y axis) recordings. In case the speed score in the bottomless car was below threshold, an open circle was used instead of the filled coloured circle (15 out of 64 in MEC (23%) and 8 out of 16 in the hippocampus (50%)). The results indicate that, although in both areas many speed cells maintain their firing properties even across extremely different contextual and behavioural situations, MEC speed cells seem to exhibit a more universal code than hippocampal speed cells. **f**, Overall distribution of running speed in the open field (o.f.; grey) and in the bottomless car trials (linear speed profile; red) for rat 14566 (Fig. 3h–j). This was the rat with the largest difference in speed across behaviours (open field, $10 \pm 8 \text{ cm s}^{-1}$; bottomless car, $20 \pm 13 \text{ cm s}^{-1}$; mean \pm s.d.). Yet the difference did not generate adaptation in the slope of the speed–rate tuning curve (Fig. 3f–h).



Extended Data Figure 9 | Theta modulation of MEC speed cells. **a**, Plots of temporal bias, as in Fig. 4a, for all MEC speed cells in the open field with weak (left) and strong (right) theta modulation as defined by the theta index (θ_{index} ; see Methods). Only the latter were prospective (discriminating threshold $\theta_{\text{index}} = 0.2$; see **b**). **b**, Temporal bias of speed cells classified according to location, task and theta modulation. Different measures are used: the maximum of the average correlation curve (peak of mean); and the mean (mean of peaks) and median (median of peaks) of the distribution of maxima of individual correlation curves. The anticipation of the speed cell response to the movements of the animal cannot be related to the learned prediction of the bottomless car protocol, since in all cases the leads are similar to those found in spontaneous open-field behaviour. Similar and even larger leads in neural activity over body kinematics have been described in the motor cortex of monkeys³⁷, as well as rats³⁸. Since the motor networks are supposed to be one of the sources of speed information feeding the hippocampal navigation systems, with prominent direct connections from secondary motor cortex to the MEC³⁹, we cannot discard the hypothesis that the lead is simply inherited from this source. Alternatively, other simple network mechanisms such as anticipated synchronization could generate this effect locally without the involvement of predictions or learning in a cognitive sense. **c**, MEC speed cells ordered according to increasing theta modulation index. Colour-coded firing rate profile across the theta cycle is plotted, with each line representing a different cell. Firing rate is normalized for visualization purposes. Red arrowheads indicate the threshold ($\theta_{\text{index}} = 0.2$) used in **a** and **b**. The plot reveals that theta-modulated cells have a characteristic behaviour, exhibiting a ramp of activity that develops roughly along the first two-thirds of the cycle and falls to near zero during the last third. **d**, Representative examples of the activity of ramping (strongly modulated, top four) and flat (weakly modulated, bottom three) speed cells at different speeds (colour-coded). Rat number is indicated in the top-left corner. Note that ramps corresponding to different speeds do not run in parallel. Instead, the ramp slope increases with speed. One

possible explanation for this is that the ramp represents the integration of speed (distance travelled) from the beginning of the theta cycle rather than speed itself. Note also that the ramp/silent division of the theta cycle roughly coincides with the reset/look-ahead division arising from the analysis of grid cell activity (Extended Data Fig. 10f, h). **e**, Normalized firing rate profile (mean \pm s.d.) for four clusters resulting from applying a *k*-means algorithm to the data in **c**. The number of clusters *k* was set to 7, and all clusters exhibiting a ramping behaviour were merged together (similar results were obtained by applying the same procedure with *k* = 4 ... 10). Note that most speed cells fall into the ramping (#1) or flat (#2) clusters. The sum of counts is 321, lower than the total cell number of 385, because 62 speed cells classified conjunctively as some other category were left out of this particular analysis and for two pure speed cells a simultaneous EEG recording was not available. **f**, Average dynamics along the theta cycle of the normalized firing rate of speed cells belonging to each of the four clusters for different running speeds (colour-coded as in **c**). **g**, First two principal components of the data. Note that the first principal component represents the ramping pattern. **h**, Scatter representation of the data in **a** across the principal components in **e**. Colours indicate clusters as in **e**. **i**, **j**, Distribution of clusters (**i**) and theta indexes (**j**) for different MEC layers. **k**, Plots obtained from the 25 most ramping (left) or flat (right) MEC speed cells (all trials). Each block shows the distribution of correlations between running speed and different temporal shifts of the instantaneous firing rate (left), together with a profile of normalized activity across the theta cycle for positive and negative acceleration with an absolute threshold of 50 cm s^{-2} (right, top) and the difference between the two curves (right, bottom; mean \pm s.e.m.). Only ramping cells express pronounced prospective behaviour, as seen both by a positive temporal shift (ramping, $206 \pm 22 \text{ ms}$, $P < 0.01$; flat, $-23 \pm 19 \text{ ms}$, $P = 0.31$; Wilcoxon signed rank tests) and by a marked difference between positive and negative acceleration curves along the ramp of activity. Friedman's tests show a significantly higher firing rate for positive acceleration in ramping cells and for negative acceleration in flat cells ($P < 0.01$).



Extended Data Figure 10 | Grid cells in MEC layer II express strong prospective theta-modulated spatial coding. **a**, Average fields of spatially modulated MEC and CA cells in bottomless car trials, filtering for only positive (red) or only negative (black) acceleration (absolute threshold, 50 cm s^{-2}). Recording layer (II, III or V) in the MEC or subfield in the hippocampus (CA1, CA3) is indicated in each case, and the average unfiltered field is shown in grey. Space is represented by the z score of the field and running direction is always defined from left to right. Note that fields were significantly shifted only in MEC layers II (strongly) and III (weakly), that is, not in MEC layer V or in the hippocampus (Fig. 4g). **b**, To rule out the possibility of a retrospective effect during negative acceleration, we restricted the analysis to the four-speed experiment. Since rats spent most of the time running at very low speed and nearly zero acceleration, the temporal bias of the average field is reduced to a minimum, and it can be used as a reliable reference. The plot shows shifted fields for different positive and negative acceleration thresholds using only data from the four-speed experiment. Acceleration threshold is colour-coded (scale bars to the right). Note that negative acceleration, regardless of its magnitude, has a very small effect on the field position, keeping the field close to the reference average field in all cases. In contrast, positive acceleration produces a prospective advance of the field that increases with acceleration threshold. **c**, Position of the average fields peaks in **b** as a function of absolute acceleration threshold when including only positive (red) or only negative (black) acceleration episodes. Note the increase in prospective shift with increasing threshold only for positive acceleration episodes. In contrast, negative acceleration produces no effect apart from a small retrospective offset. Such an offset is expected as a consequence of prospection during positive acceleration, since the average field at the lowest speed, used as a reference, should have a small, yet non-zero, prospective bias. **d**, Shifted fields as in **b**, but using only cells that could be classified as grid cells based on rotational symmetry in a complementary open field recording (using the 99th percentile of a shuffled distribution as the classification criterion). The absolute acceleration threshold was 50 cm s^{-2} . **e**, Shifts that maximized the correlation

between positive or negative acceleration-related fields and the reference average field shown in **d** (mean \pm s.e.m.; *Wilcoxon signed-rank test after Holms–Bonferroni correction, $P < 0.01$). **f**, Phase map of the pool of all putative grid cells, indicating ‘look ahead’ and ‘reset’ stages over two theta cycles (see **h**). In the look ahead stage, the grid network engages in forward sweeps, related to phase precession proper²². In the reset stage the spatial representation suffers a sudden jump back, opposite to the running direction, and the correlation between grid cell firing phase and position is very poor. **g**, Similar phase maps filtering for only positive (top) or only negative (bottom) acceleration (absolute acceleration threshold, 50 cm s^{-2}). **h**, Top: average firing rate along two theta cycles. The local minima, indicated with dashed lines, were used to define the frontiers between the look ahead and reset stages^{7,32,40,41}. During the look ahead stage, phase precession proper takes place, while during the reset stage, the spatial code jumps back and remains relatively static as theta phase increases (see **f**). Middle, in three consecutive rows, the average dynamics of Δ along two theta cycles for different acceleration thresholds (colour-coded; from top to bottom: MEC layers II, III and V). Note that the prospective shift of grid fields increases during the reset stage and decreases during the look ahead stage. This speaks strongly against the idea that the prospective effect is a by-product of forward sweeps of different magnitude, and in favour of transient and local distortions in the representation of location. Bottom, acceleration is not strongly modulated by theta phase, as observed when computing the overall average (grey) and the average restricted to positive (red) or negative (black) acceleration. **i**, Frequency distribution of ratio between intrinsic firing frequency and local field potential (LFP) theta frequency in grid cells and speed cells. In grid cells (green), the mean intrinsic firing frequency is 3% higher than the theta frequency obtained from the LFP power spectrum (Mann–Whitney U -test, $P = 1 \times 10^{-21}$). This difference is due to phase precession. In contrast, in speed cells (grey), the mean intrinsic firing frequency is only 0.6% higher than the LFP theta frequency ($P = 0.043$), suggesting that a similar mechanism is not present in this population.

Crystal structures of a polypeptide processing and secretion transporter

David Yin-wei Lin^{1,2}, Shuo Huang^{2†} & Jue Chen^{1,2}

Bacteria secrete peptides and proteins to communicate, to poison competitors, and to manipulate host cells. Among the various protein-translocation machineries, the peptidase-containing ATP-binding cassette transporters (PCATs) are appealingly simple. Each PCAT contains two peptidase domains that cleave the secretion signal from the substrate, two transmembrane domains that form a translocation pathway, and two nucleotide-binding domains that hydrolyse ATP. In Gram-positive bacteria, PCATs function both as maturation proteases and exporters for quorum-sensing or antimicrobial polypeptides. In Gram-negative bacteria, PCATs interact with two other membrane proteins to form the type 1 secretion system. Here we present crystal structures of PCAT1 from *Clostridium thermocellum* in two different conformations. These structures, accompanied by biochemical data, show that the translocation pathway is a large α -helical barrel sufficient to accommodate small folded proteins. ATP binding alternates access to the transmembrane pathway and also regulates the protease activity, thereby coupling substrate processing to translocation.

Protein translocation is a complex yet essential task for life. The general secretion pathway, Sec translocon, transports many secreted and plasma membrane proteins across the eukaryotic endoplasmic reticulum or bacterial plasma membranes. In addition, prokaryotes use several dedicated transport systems to secrete specific cargo proteins for the benefit of survival. The simplest Sec-independent pathway is a dual-function ATP-binding cassette (ABC) transporter that processes and secretes polypeptides¹. ABC transporters are a large family of membrane proteins that harness the energy from ATP hydrolysis to drive substrate translocation^{2,3}. All ABC transporters share a conserved architecture of two transmembrane domains (TMDs) and two nucleotide-binding domains (NBDs). The transporters involved in protein secretion are unique among ABC transporters, as they contain additional peptidase domains essential for substrate processing (Fig. 1a). These peptidase domains belong to the cysteine protease superfamily, classified as family C39, bacteriocin-processing peptidase⁴.

In Gram-positive bacteria, the PCATs are responsible for exporting quorum-sensing or antimicrobial peptides called bacteriocins⁵. Substrates of PCATs are synthesized as precursors with an amino-terminal leader peptide containing the consensus sequence L(−12)XXXE(−8)L(−7)XXXXG(−2)G(−1). Secretion of the cargo peptide requires proteolytic cleavage of the leader peptide at the conserved double-glycine motif.

In Gram-negative bacteria, PCATs interact with an outer membrane protein and another accessory protein to form a type 1 secretion system (T1SS) that secretes proteins from the cytosol directly into the extracellular space⁶. The substrates secreted by the T1SS range from small antibiotic peptides (microcins) to large adhesion proteins of 900 kilodaltons (kDa). Protein substrates larger than 10 kDa usually contain a carboxy-terminal secretion signal that is not subject to proteolytic processing. The peptidase domains of the corresponding PCATs have no enzymatic activity owing to the absence of a catalytic cysteine residue. Nevertheless, these degenerate peptidase domains are essential in recruiting the substrate to the transporter⁷.

So far, structural studies of PCATs are limited to isolated peptidase domains and NBDs^{7–12}. Here, we report the crystal structure of a full-length PCAT in two different conformations. These structures, correlated with functional data, support a model for understanding how small protein substrates are processed and secreted by dual-functioning ABC transporters.

Biochemical characterization

To enable structural studies of PCATs, we screened the expression and purification of approximately 50 members of this family from various prokaryotic species. Crystals were obtained from a *C. thermocellum* transporter (PCAT1) whose sequence is 55% identical to that of LagD, a bacteriocin transporter from *Lactococcus lactis*¹, and 30% identical to that of HlyB, a T1SS transporter in *Escherichia coli* that secretes a 1,024-residue toxin^{13,14} (Extended Data Fig. 1).

The substrate of PCAT1 has not been characterized. In the operon downstream from PCAT1, one gene encodes a 90-residue protein with the N-terminal consensus leader peptide (Fig. 1b). As PCATs and their corresponding substrates are usually encoded in the same gene cluster, we tested whether the product of this gene was subject to enzymatic cleavage by PCAT1. Proteoliposomes containing wild-type PCAT1 converted the putative substrate into a product approximately 2,500 Da smaller, consistent with removal of the double-glycine leader peptide (Fig. 1c). The active site of C39 peptidase consists of a cysteine and a histidine. Cleavage was eliminated when the corresponding catalytic residues in PCAT1 (Cys 21 or His 99) were mutated to alanine (Fig. 1c). No protease activity was observed when PCAT1 was incubated with known substrates of closely related transporters in the PCAT subfamily (Extended Data Fig. 2).

Next we tested whether the function of the peptidase domains is linked to the function of the nucleotide-binding domains. When the recombinant substrate was incubated with proteoliposomes containing PCAT1, robust cleavage was observed only when ATP was not bound to the transporter. Specifically, cleavage occurred in the absence of nucleotide and in the presence of ADP (Fig. 1d). The only condition under which robust cleavage occurred in the presence of

¹Laboratory of Membrane Biology and Biophysics, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA. ²Howard Hughes Medical Institute, 1230 York Avenue, New York, New York 10065, USA. [†]Present address: School of Chemistry and Biochemistry, Georgia Institute of Technology, 901 Atlantic Drive, Atlanta, Georgia 30332, USA.

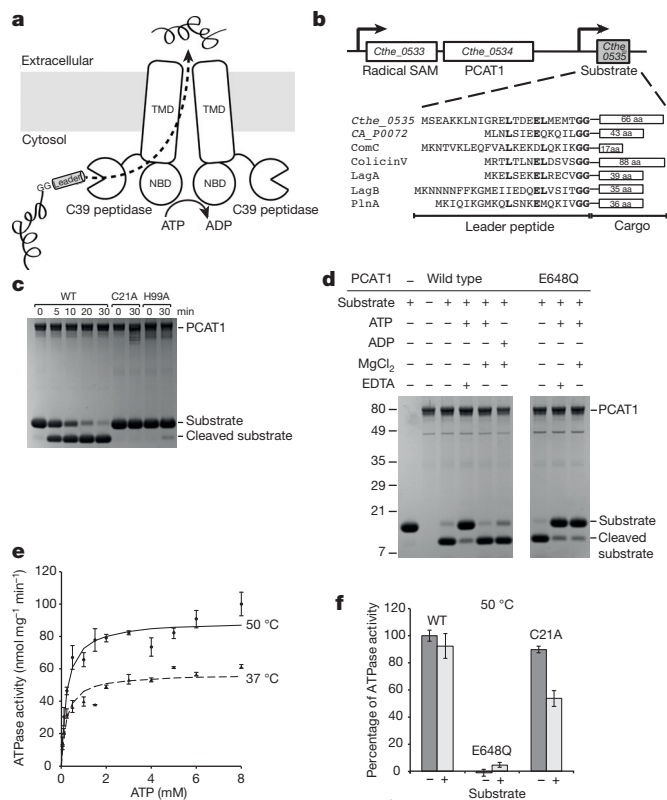


Figure 1 | Biochemical properties of PCAT1. **a**, The domain structure of PCAT. **b**, The genetic organization of *C. thermocellum*. Gene *Cthe_0533* encodes a radical SAM enzyme, *Cthe_0534* encodes PCAT1, and *Cthe_0535* encodes the substrate. Also shown is the sequence alignment of the N-terminal leader peptides of different PCAT substrates. aa, amino acids. **c**, The protease activity of wild-type (WT) and mutant PCAT1 in proteoliposomes measured at 50 °C. **d**, Substrate cleavage is inhibited by ATP but not ADP. Molecular weights in kDa. **e**, The ATPase activity of PCAT1 in proteoliposomes. Data points represent the means and standard deviations of 3–9 measurements. The maximum activity (V_{\max}) at 37 °C and 50 °C, determined by nonlinear regression of the Michaelis–Menten equation, are 57 nmol mg⁻¹ min⁻¹ (10 per minute) and 90 nmol mg⁻¹ min⁻¹ (20 per minute), respectively. **f**, ATPase activity in the presence and absence of substrate.

ATP was when Mg²⁺ was also present, which mediates hydrolysis of ATP to ADP (Fig. 1d). By contrast, ATP inhibited substrate cleavage in the absence of Mg²⁺ or when a mutation was introduced into the NBD (Fig. 1d). This mutation (E648Q) permits ATP binding but prevents hydrolysis^{15,16}. Thus, the data show that ATP binding to the NBDs inhibits substrate cleavage.

C. thermocellum is a thermophilic bacterium that grows in a large range of temperatures¹⁷. At the two temperatures we tested, 37 °C and 50 °C, PCAT1 hydrolyses ATP with a Michaelis constant (K_m) for ATP of 0.23 mM, typical for ABC transporters (Fig. 1e). The maximum ATP turnover rate at 50 °C is approximately 20 per minute, twice the 37 °C rate (Fig. 1e). A broad range of turnover capacity has been observed across the family of ABC transporters. For example, the transporter associated with antigen processing (TAP) and the maltose transporter are on the high end of the spectrum with rates approaching 300–400 per minute^{18,19}, whereas the cystic fibrosis transmembrane conductance regulator (CFTR), an ATP-gated ion channel, is on the low end with rates similar to PCAT1 (refs 20, 21). ATP hydrolysis by ABC transporters is generally stimulated by the presence of substrate. We did not observe this behaviour in PCAT1 (Fig. 1f). In fact, instead of stimulation we observed a small but reproducible reduction in hydrolysis rate upon substrate addition in both the

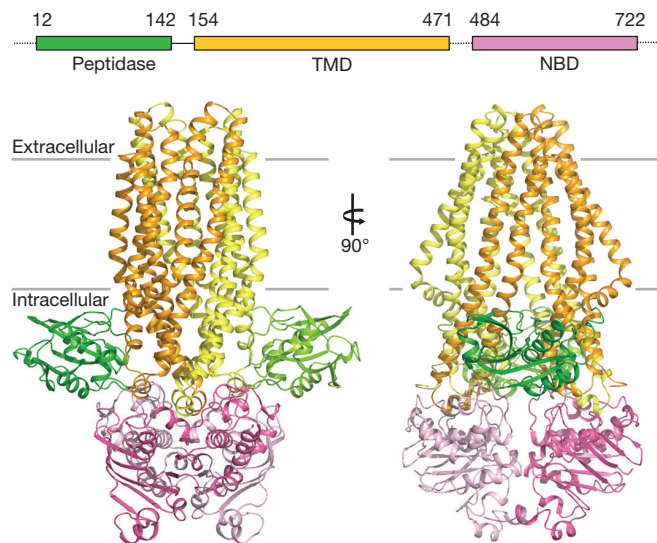


Figure 2 | Ribbon diagram of the structure of PCAT1. Green, peptidase domain; yellow, TMD; magenta, NBD. The two subunits are distinguished by different shades. The residue numbers of each domain are indicated in the schematic drawing.

wild-type protein and the cleavage-incompetent mutant C21A (Fig. 1f).

These biochemical data suggest that PCAT1 processes and secretes a small protein of 66 residues, generated after removal of the leader peptide. The protease activity of PCAT1 appears to be specific and is dependent upon the catalytic cysteine and histidine residues. Binding of ATP inhibits the peptidase activity. The ATPase activity of PCAT1 is low and rather insensitive to the presence of substrate. Thus, we seek to understand how the structure of PCAT1 can explain these biochemical properties.

Structural determination of the ATP-free form

Using the lipidic bicelle method²², we obtained crystals of PCAT1 in two space groups: $P2_12_12_1$, which diffracted X-rays to 3.6 Å, and $C222_1$, which diffracted to 4.1 Å (Extended Data Table 1). In both crystal forms, the initial phase was obtained by molecular replacement using the TMD and NBD of the ABC exporter TM287/288 (Protein Data Bank (PDB) accession 3QF4) and the peptidase domain of *Streptococcus* ComA (PDB accession 3K8U) as search models. The asymmetric unit of the $P2_12_12_1$ crystal system consists of two monomers that form a dimer, whereas that of the $C222_1$ crystal contains only one monomer. A dimer of PCAT1 was generated by the crystallographic two-fold symmetry. Despite their different crystal packing arrangements, the structures of the PCAT1 dimer are essentially identical in the two crystal forms. The higher-resolution structure obtained in $P2_12_12_1$ was further refined and used to generate the figures in this paper.

For unambiguous assignment of the amino acid sequence, we crystallized the native protein and three methionine-substitution mutants with selenomethionine labels. Anomalous difference analysis at the selenium edge identified the positions of 19 native and the 9 additionally introduced selenomethionine residues, providing register markers on every transmembrane (TM) helix (Extended Data Fig. 3). The final model, containing 695 out of 727 residues, was refined to an R_{work} of 25% and an R_{free} of 29% with good geometry (Extended Data Fig. 1).

The structure of PCAT1 in the absence of ATP

PCAT1 is a symmetrical dimer; each monomer contains an N-terminal C39 peptidase domain, a TMD of six TM helices, and a C-terminal NBD (Fig. 2). The overall architecture of the TMDs and

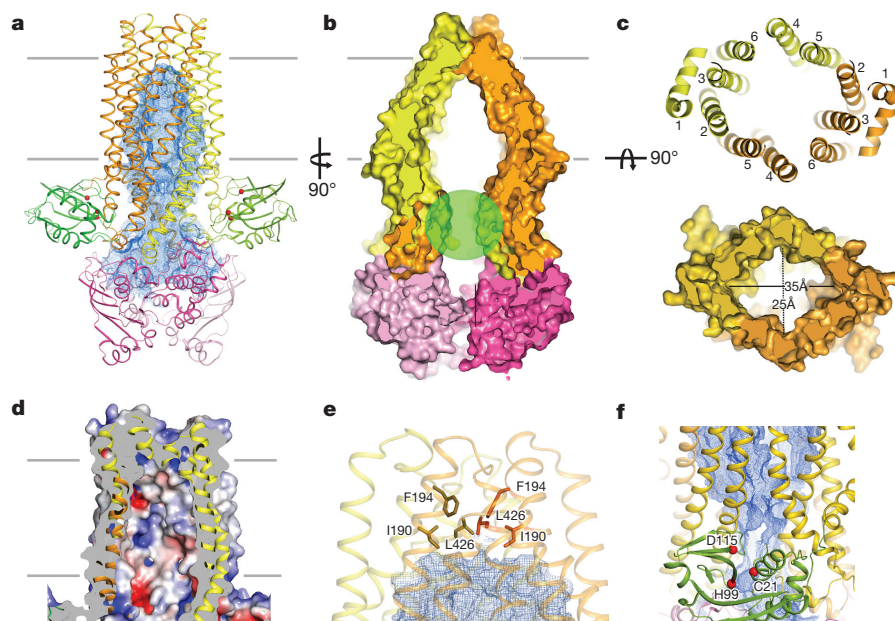


Figure 3 | The translocation pathway. **a**, The large TM tunnel, shown as a blue mesh. The three catalytic residues in the peptidase domains are indicated by red balls. **b**, A slab view of PCAT1. The green circle indicates the surface where the peptidase domain binds. **c**, The cross-section of the TM tunnel in a ribbon diagram (top) and a surface view (bottom). The TM helices are labelled.

NBDs are similar to those of other ABC exporters. The TMDs extend into the cytosol, placing the NBDs away from the membrane. The two NBDs form a semi-open dimer, separated at the TMD–NBD interfaces and making contacts at the distal end of the structure. The peptidase domains, observed for the first time in an ABC transporter, are positioned at two opposite sides of the transporter, making contact with both the TMDs and NBDs (Fig. 2).

The translocation pathway in PCAT1 is a large α -helical barrel traversing nearly the entire lipid bilayer (Fig. 3a, b). The cross-section of the barrel is rhomboidal and has an area of approximately 440 \AA^2 in the membrane-spanning region (Fig. 3c). The interior surface of the TM tunnel is lined with charged residues, providing a hydrophilic environment for the cargo protein (Fig. 3d). Near the extracellular surface, hydrophobic residues I190, F194 and L426 of both subunits make van der Waals contacts with each other to form a closed gate (Fig. 3e). In the cytoplasm, the helical barrel tapers down to an interior diameter of 10–12 \AA and opens laterally to the cytosol on both sides of the molecule.

The peptidase domains dock onto the lateral openings of the TM tunnel, with the catalytic site facing the gateway (Fig. 3f). They appear to be positioned perfectly to process the substrate at its leader peptide by enzymatic cleavage, while recruiting the substrate into the translocation pathway for secretion. The buried surface of the peptidase domain at the interface is relatively small (approximately 980 \AA^2), which suggests a weak association between the peptidase domain and the rest of the transporter. The structure of the peptidase domain is very similar to those of isolated ComA⁸ and HlyB⁷ (root mean squared deviation (r.m.s.d) of 1.9 and 2.1 \AA , respectively), indicating that interactions with the rest of the transporter do not induce any structural changes in the peptidase domain.

The configuration of the NBD dimer in PCAT1 resembles that of the pre-translocation state of the maltose transporter, a substrate-induced conformation primed for ATP hydrolysis²³. One of the key features of this conformation is that two highly conserved ATP-binding residues, the Walker A serine and the switch histidine, are located at the dimer interface and make contact with the conserved D-loop of the opposite NBD (Fig. 4). In this configuration, ATP binding would

complete the transition to the hydrolysis-competent state, where residues from both NBDs interact with the γ -phosphate of ATP to form a closed dimer²³. By contrast, in the resting state of the maltose transporter, the two NBDs make no contact with each other and ATP

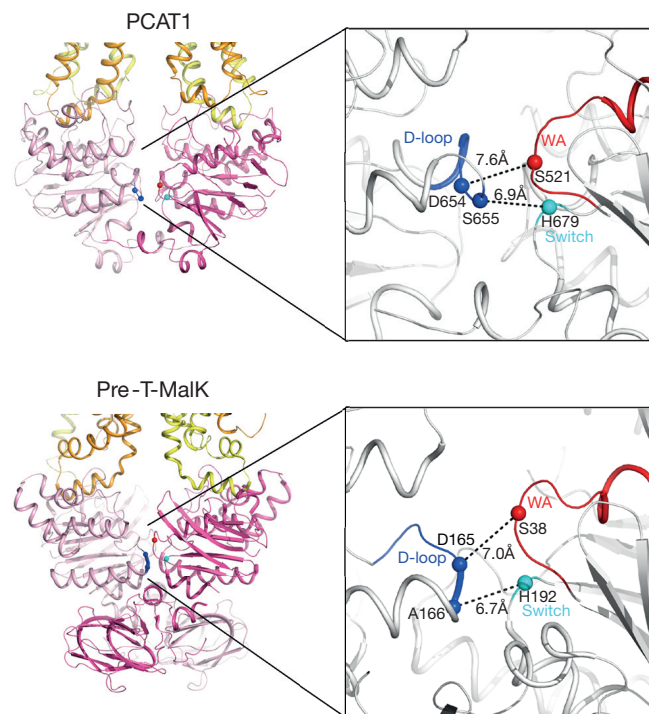


Figure 4 | A primed NBD dimer. Comparison of the NBD dimer in PCAT1 and the pre-T state of the maltose transporter (MalK). Key residues at the dimer interface are indicated as balls. Red, Walker A (WA); blue, D-loop; cyan, switch histidine. The distances between equivalent residues in pre-T-MalK (PDB accession 3PV0) and PCAT1 are indicated. In contrast, in the MalK resting state, where the NBDs are open, the distances between D165/S38 and A166/H192 are 10.3 \AA and 10.7 \AA , respectively.

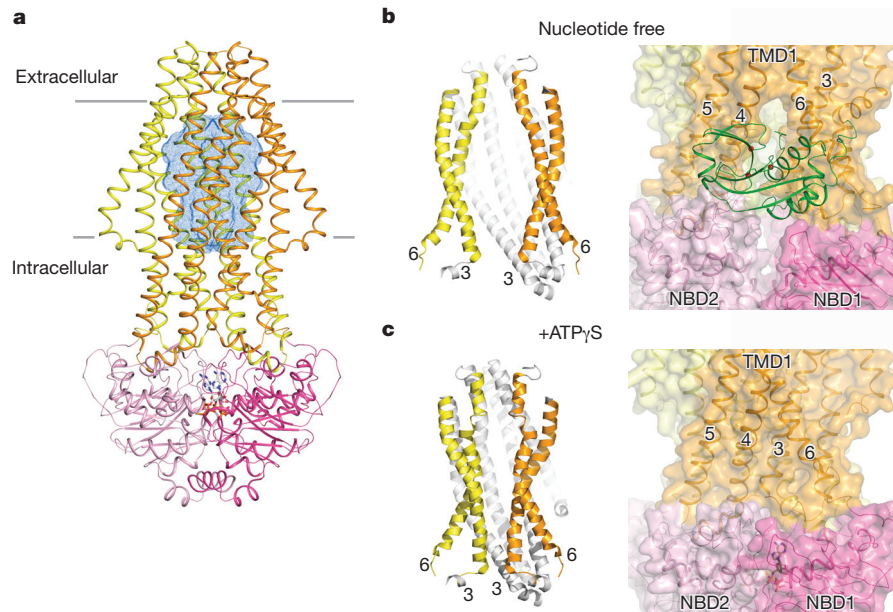


Figure 5 | Conformational change upon ATP binding. **a**, Ribbon diagram of the ATP γ S-bound form. The TM cavity is presented as a blue mesh. **b**, In the ATP-free form, the peptidase domain docks onto the cytoplasmic opening. **c**, Closure of the cytoplasmic opening in the ATP-bound state.

binding does not promote the global conformational changes necessary for hydrolysis²⁴. The pre-translocation state of the maltose transporter is induced by the substrate-loaded binding protein, a mechanism that couples the substrate to ATP hydrolysis²³. This primed NBD dimer of PCAT1, obtained in the absence of the cargo protein, is consistent with biochemical data showing that the cargo peptide does not stimulate ATP hydrolysis (Fig. 1f).

Structure of PCAT1 bound with ATP

Next, we determined the crystal structure of the E648Q mutant in complex with ATP γ S. The electron density map (Extended Data Fig. 4), calculated at a resolution of 5.5 Å, reveals an overall conformation that is very different from that of the ATP-free form (Fig. 5a). The final model, refined with deformable elastic network restraints to an R_{work} of 30% and R_{free} of 31%, does not contain the peptidase

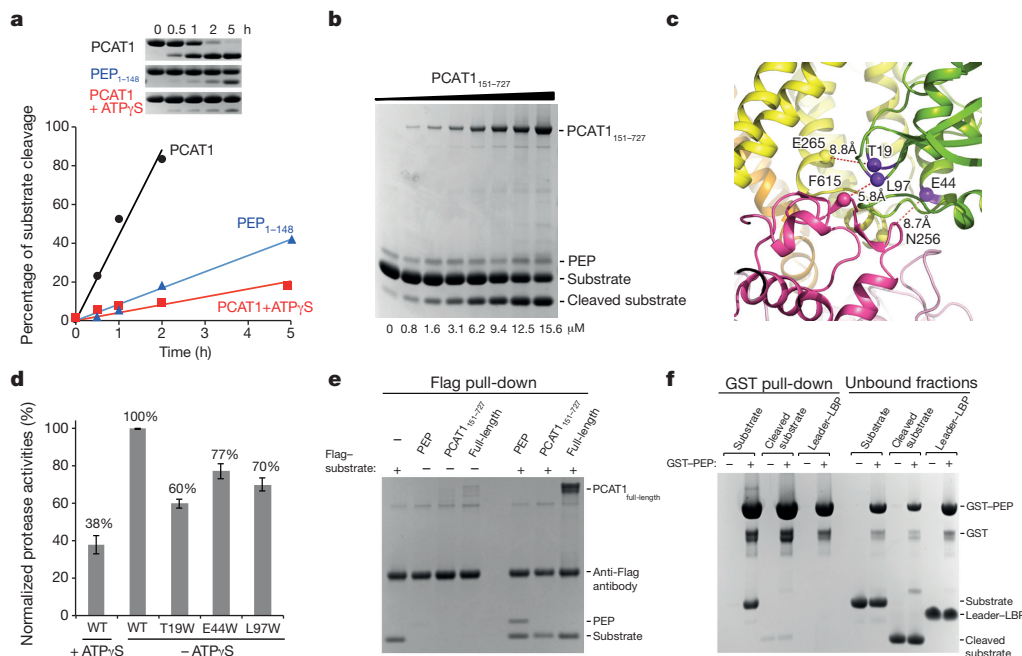


Figure 6 | Functional properties of the PEP domain. **a**, Time course of substrate cleavage at 37 °C. **b**, Trans-complementation assay. Substrate cleavage was monitored over a 2 h time course at 37 °C in reaction mixtures containing 3 μ M PEP, 77 μ M substrate, and a series of PCAT1_{151–727} concentrations. **c**, The interface between the peptidase domain (green), TMD (yellow) and NBD (magenta) in the inward-facing conformation. The C α distances of T19, E44 and L97 from their interacting residues are indicated.

d, The protease activities in the *trans*-complementation assay, normalized on the basis of the activity of the wild-type PEP in the absence of nucleotide. **e**, Pull-down assay using an antibody-conjugated resin against the Flag-tagged substrate. **f**, Pull-down assay using the glutathione resin against the GST-tagged PEP construct. Leader-LBP: a construct containing the leader peptide fused to a lanthanide-binding peptide.

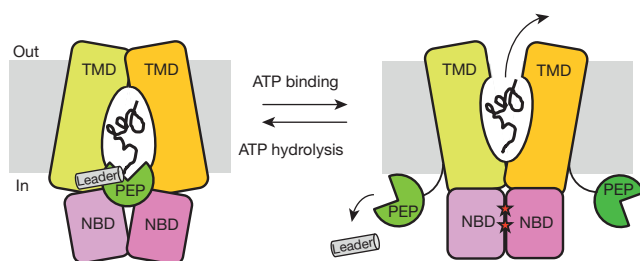


Figure 7 | The alternating-access model for protein translocation. In the absence of ATP, the substrate is recruited to the transporter through the peptidase domain and inserted into the translocation pathway. Proteolytic cleavage takes place in this conformation to remove the leader peptide. ATP binding alters the access of the translocation pathway and disengages the peptidase domains. ATP hydrolysis resets the transporter to the inward-facing conformation.

domains, as no electron density was observed for these domains (Extended Data Table 1).

In the ATP-free form, the peptidase domain interacts with a surface where a TM tunnel opens and the NBDs are separated, making approximately 60% of the contacts with TM helices 3 and 6 on one side of the opening and 40% of contacts with TM helix 4 and an NBD on the other side of the opening (Fig. 5b). This surface is completely changed upon ATP binding. The two NBDs rotate inward to form a closed dimer and TM helices 3–6 shift towards the molecular centre. Thus, the TM pathway is narrowed and the lateral openings are closed (Fig. 5c). The extracellular gate remains closed, resulting in an occluded cavity separated from both sides of the membrane (Fig. 5a). In this conformation, the peptidase domains can no longer bind to the same region, and instead are flexibly attached to the rest of the molecule through the covalent linkers.

Correlation of structure with function

Can the dissociation of the peptidase domains account for the lower proteolytic activity upon ATP binding (Fig. 1d)? To address this question, we expressed and purified the peptidase domain in isolation (PEP) and compared its activity with the full-length protein (Fig. 6a). Previously, the catalytic activity of isolated peptidase domains have been reported for LagD¹, CvaB²⁵ and ComA²⁶. In those systems, however, full-length transporters were not studied for comparison. Here we show that at the same enzyme/substrate ratio, isolated PEP cleaves the substrate 80% more slowly than the nucleotide-free full-length PCAT1. Addition of ATP γ S reduces the cleavage rate of the full-length PCAT1 by 90%, a level comparable with isolated PEP. The protease activity of the ATP γ S-bound form is therefore consistent with the crystal structure, which shows that ATP γ S causes the peptidase domains to dissociate and thus become isolated from the rest of the transporter.

To test further if association of the peptidase domain with the rest of the transporter is necessary for efficient substrate cleavage, we carried out a *trans*-complementation assay in which isolated PEP was added *in trans* to a construct containing only the TMD and NBD (PCAT1_{151–727}). Figure 6b shows that at a constant PEP concentration, substrate cleavage increased as the concentration of PCAT1_{151–727} increased, probably due to the formation of a PEP–PCAT1_{151–727} complex mimicking the full-length protein. Consistent with the crystal structures, the presence of ATP γ S reduced the *trans*-complementation efficiency to 38% and point mutations at the PEP-docking interface reduced substrate cleavage to 60–77% of the wild-type protein (Fig. 6c, d).

To test whether substrate binding requires the full-length transporter, we carried out substrate pull-down experiments using isolated PEP, full-length PCAT1, and the PEP-truncated construct PCAT1_{151–727} (Fig. 6e, f). It is clear that the peptidase domain is

necessary and sufficient for substrate binding (Fig. 6e, f). As all three catalytic residues reside in the peptidase domain, the most likely explanation for the higher protease activity of the full-length protein is that the TMDs have a role in orientating the substrate for optimal cleavage, reminiscent of the function of the LSGGQ loop in positioning ATP for hydrolysis²³.

Using the pull-down assay, we further showed that only the uncleaved substrate binds tightly to PEP (Fig. 6f). Two other constructs, the mature substrate without the leader peptide and a pseudo-substrate consisting of the leader peptide fused to a lanthanide-binding peptide, failed to pull down with the glutathione S-transferase (GST)-tagged PEP (Fig. 6f). These data indicate that the cleavage products have lower affinity for PEP. After cleavage, the mature substrate would be readily released for translocation and the leader peptide would be exchanged by an uncleaved substrate to initiate a new transport cycle.

Discussion

We characterized the structure and function of an ABC transporter that processes and secretes a 66-residue protein. The protein-conducting pathway observed here is fundamentally different from that of the Sec translocon. In the Sec translocon, the TM channel is constricted in the middle of the membrane²⁷. The size of the pore at the constriction point would only allow passage of extended polypeptide chains or a single α -helix^{28,29}. Furthermore, the TM pathway of the Sec translocon contains a lateral gateway to the lipid bilayer for the insertion of hydrophobic peptides^{30,31}. In contrast, the TM tunnel of PCAT1 is completely shielded from the membrane and the cavity is large enough to accommodate a small folded domain (Extended Data Fig. 5). These structural differences in the two protein-transporting systems correlate well with their distinct functional requirements. The Sec translocon inserts hydrophobic peptides into the membrane and transfers hydrophilic peptides across it. The peptide in transit is probably in an extended conformation or possibly a compact α -helix^{32,33}. PCATs, on the other hand, only secrete polypeptides across the membrane into the extracellular medium. The large size and the hydrophilic interior of the TM pathway suggest that PCAT can allow at least partial folding of the cargo protein while it is being translocated across the membrane.

The prevailing model for active transport is alternating access, in which the TM pathway is alternately exposed to one side of the membrane. This principle holds for all ABC transporters that have been studied so far. The question is whether this model can also apply to protein translocation. In other words, whether PCATs are ATP-driven transporters, in which case alternate access would be a strict requirement, or ATP-gated channels akin to CFTR.

Given the structural and functional data presented here, we can envision a model consistent with the classic alternating-access mechanism (Fig. 7). In the absence of ATP, PCAT1 reveals an inward-facing conformation, in which a large translocation pathway is open to the cytosol at the very site where the peptidase domain is docked. We also know that PCAT1 is proteolytically active in the absence of ATP. It is possible that in this conformation, the C-terminal region of the substrate could snuggle into the translocation pathway whereas the N-terminal leader peptide remains associated with the peptidase domain. Such orientation of the substrate may be optimal for proteolytic cleavage to free the cargo from the leader peptide. Once cleavage has taken place, ATP binding would induce closure of the cytoplasmic opening and dissociation of the peptidase domains. The crystal structure of the ATP γ S-bound form shows an occluded cavity. However, we imagine that if the substrate were inside the cavity it might favour opening of the extracellular gate, resulting in an outward-facing conformation to release the substrate. ATP hydrolysis then resets the transporter to the inward-facing conformation. The lower protease activity in the ATP-bound form, in which the peptidase domains are disengaged from the translocation pathway, prevents cleavage of the

substrate when it is not in position for translocation. While this model could apply nicely to PCATs that secrete small proteins, it is interesting to consider that some PCATs transport very large proteins, up to 900 kDa. It seems only possible that those PCATs function like the channel versions of ABC transporters.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 December 2014; accepted 8 June 2015.

- Håvarstein, L. S., Diep, D. B. & Nes, I. F. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. *Mol. Microbiol.* **16**, 229–240 (1995).
- ter Beek, J., Guskov, A. & Slotboom, D. J. Structural diversity of ABC transporters. *J. Gen. Physiol.* **143**, 419–435 (2014).
- Rice, A. J., Park, A. & Pinkett, H. W. Diversity in ABC transporters: type I, II and III importers. *Crit. Rev. Biochem. Mol. Biol.* **49**, 426–437 (2014).
- Rawlings, N. D., Waller, M., Barrett, A. J. & Bateman, A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **42**, D503–D509 (2014).
- Gebhard, S. ABC transporters of antimicrobial peptides in Firmicutes bacteria - phylogeny, function and regulation. *Mol. Microbiol.* **86**, 1295–1317 (2012).
- Lenders, M. H., Reimann, S., Smits, S. H. & Schmitt, L. Molecular insights into type I secretion systems. *Biol. Chem.* **394**, 1371–1385 (2013).
- Lecher, J. et al. An RTX transporter tethers its unfolded substrate during secretion via a unique N-terminal domain. *Structure* **20**, 1778–1787 (2012).
- Ishii, S. et al. Crystal structure of the peptidase domain of *Streptococcus* ComA, a bifunctional ATP-binding cassette transporter involved in the quorum-sensing pathway. *J. Biol. Chem.* **285**, 10777–10785 (2010).
- Ishii, S., Yano, T., Okamoto, A., Murakawa, T. & Hayashi, H. Boundary of the nucleotide-binding domain of *Streptococcus* ComA based on functional and structural analysis. *Biochemistry* **52**, 2545–2555 (2013).
- Zaitseva, J. et al. A structural analysis of asymmetry required for catalytic activity of an ABC-ATPase domain dimer. *EMBO J.* **25**, 3432–3443 (2006).
- Schmitt, L., Benabdelhak, H., Blight, M. A., Holland, I. B. & Stubbs, M. T. Crystal structure of the nucleotide-binding domain of the ABC-transporter haemolysin B: identification of a variable region within ABC helical domains. *J. Mol. Biol.* **330**, 333–342 (2003).
- Zaitseva, J., Jenewein, S., Jumpertz, T., Holland, I. B. & Schmitt, L. H662 is the linchpin of ATP hydrolysis in the nucleotide-binding domain of the ABC transporter HlyB. *EMBO J.* **24**, 1901–1910 (2005).
- Gentschev, I. & Goebel, W. Topological and functional studies on HlyB of *Escherichia coli*. *Mol. Gen. Genet.* **232**, 40–48 (1992).
- Koronakis, V. & Hughes, C. Bacterial signal peptide-independent protein export: HlyB-directed secretion of hemolysin. *Semin. Cell Biol.* **4**, 7–15 (1993).
- Orelle, C., Dalmás, O., Gros, P., Di Pietro, A. & Jault, J. M. The conserved glutamate residue adjacent to the Walker-B motif is the catalytic base for ATP hydrolysis in the ATP-binding cassette transporter BmrA. *J. Biol. Chem.* **278**, 47002–47008 (2003).
- Moody, J. E., Millen, L., Binns, D., Hunt, J. F. & Thomas, P. J. Cooperative, ATP-dependent association of the nucleotide binding cassettes during the catalytic cycle of ATP-binding cassette transporters. *J. Biol. Chem.* **277**, 21111–21114 (2002).
- Sissons, C. H., Sharrock, K. R., Daniel, R. M. & Morgan, H. W. Isolation of cellulolytic anaerobic extreme thermophiles from New Zealand thermal sites. *Appl. Environ. Microbiol.* **53**, 832–838 (1987).
- Gorbulev, S., Abele, R. & Tampe, R. Allosteric crosstalk between peptide-binding, transport, and ATP hydrolysis of the ABC transporter TAP. *Proc. Natl Acad. Sci. USA* **98**, 3732–3737 (2001).
- Davidson, A. L., Shuman, H. A. & Nikaido, H. Mechanism of maltose transport in *Escherichia coli*: transmembrane signaling by periplasmic binding proteins. *Proc. Natl Acad. Sci. USA* **89**, 2360–2364 (1992).
- Rosenberg, M. F., Kamis, A. B., Aleksandrov, L. A., Ford, R. C. & Riordan, J. R. Purification and crystallization of the cystic fibrosis transmembrane conductance regulator (CFTR). *J. Biol. Chem.* **279**, 39051–39057 (2004).
- Li, C. et al. ATPase activity of the cystic fibrosis transmembrane conductance regulator. *J. Biol. Chem.* **271**, 28463–28468 (1996).
- Ujwal, R. & Abramson, J. High-throughput crystallization of membrane proteins using the lipidic bicelle method. *J. Vis. Exp.* **59**, e3383 (2012).
- Oldham, M. L. & Chen, J. Crystal structure of the maltose transporter in a pretranslocation intermediate state. *Science* **332**, 1202–1205 (2011).
- Khare, D., Oldham, M. L., Orelle, C., Davidson, A. L. & Chen, J. Alternating access in maltose transporter mediated by rigid-body rotations. *Mol. Cell* **33**, 528–536 (2009).
- Wu, K. H. & Tai, P. C. Cys32 and His105 are the critical residues for the calcium-dependent cysteine proteolytic activity of CvaB, an ATP-binding cassette transporter. *J. Biol. Chem.* **279**, 901–909 (2004).
- Ishii, S., Yano, T. & Hayashi, H. Expression and characterization of the peptidase domain of *Streptococcus pneumoniae* ComA, a bifunctional ATP-binding cassette transporter involved in quorum sensing pathway. *J. Biol. Chem.* **281**, 4726–4731 (2006).
- van den Berg, B. et al. X-ray structure of a protein-conducting channel. *Nature* **427**, 36–44 (2004).
- Park, E. et al. Structure of the SecY channel during initiation of protein translocation. *Nature* **506**, 102–106 (2014).
- Gogala, M. et al. Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion. *Nature* **506**, 107–110 (2014).
- Zimmer, J., Nam, Y. & Rapoport, T. A. Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* **455**, 936–943 (2008).
- Egea, P. F. & Stroud, R. M. Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. *Proc. Natl Acad. Sci. USA* **107**, 17182–17187 (2010).
- Mingarro, I., Nilsson, I., Whitley, P. & von Heijne, G. Different conformations of nascent polypeptides during translocation across the ER membrane. *BMC Cell Biol.* **1**, 3 (2000).
- Kowarik, M., Kung, S., Martoglio, B. & Helenius, A. Protein folding during cotranslational translocation in the endoplasmic reticulum. *Mol. Cell* **10**, 769–778 (2002).

Acknowledgements We thank the staff at the Advance Photon Source GM/CA-CAT and NE-CAT for assistance with data collection, S. McCarry for editing the manuscript, R. MacKinnon and D. Kearns for helpful discussions, M. L. Oldham for help with figure preparation, and H. Zhang and W. Mi for efforts in the early stage of this project. This work was supported by the Howard Hughes Medical Institute.

Author Contributions All authors designed the study and analysed the data. D.Y.-w.L. and S. H. performed cloning and biochemical experiments. D.Y.-w.L. determined the crystal structures and wrote the manuscript together with J.C.

Author Information Coordinates and structure factors have been deposited in the Protein Data Bank under accession numbers 4RY2 (nucleotide-free form) and 4SOF (ATPγS-bound form). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.C. (juechen@rockefeller.edu).

METHODS

No statistical methods were used to predetermine sample size.

Protein expression and purification. Full-length PCAT1 was cloned into vector pMCSG20 with an N-terminal GST tag from the genomic DNA of *C. thermocellum* (ATCC 27405). PCAT1-containing plasmids were transformed into *E. coli* BL21(DE3) codon plus (RIL) cells. Cells were grown in Terrific Broth (Novagen) supplemented with ampicillin at 20 °C until OD_{600 nm} = 0.6–0.8 and induced with 100 µM isopropyl-β-D-thiogalactoside (IPTG) at 16 °C for 24 h before collection by centrifugation at 4,000g for 15 min. Cells were resuspended in a lysis buffer containing 50 mM Tris-HCl pH 7.0, 500 mM NaCl, and 10% glycerol, lysed by two passes through a high-pressure homogenizer (Emulsiflex-C3; Avestin), and centrifuged at 80,000g for 40 min to isolate the membrane fraction. Harvested membranes were stored at –80 °C. For purification, membranes were solubilized in the lysis buffer with the addition of 1% *n*-dodecyl β-D-maltoside (DDM) and 5 mM dithiothreitol (DTT) at 4 °C for 2 h and spun at 70,000g for 20 min to remove the insoluble fraction. Solubilized membranes were incubated with Glutathione Sepharose 4B resins (GE Healthcare) and washed extensively with the lysis buffer with 2 mM *n*-undecyl-β-D-maltopyranoside (UDM) and 5 mM DTT. Tobacco etch virus (TEV) protease was added to PCAT1-bound resins and incubated overnight at 4 °C. Cleaved PCAT1 proteins were eluted, concentrated, and further purified by gel-filtration chromatography (Superdex 200 16/60) in a buffer containing 20 mM Tris-HCl pH 7.0, 150 mM NaCl, 2 mM UDM and 5 mM DTT.

Selenomethionine-incorporated proteins were generated by expressing PCAT1 in B834 (DE3) cells in M9 minimal media supplemented with glucose, vitamins, and amino acids, with the exception of L-methionine (Molecular Dimensions). There are 21 native methionine residues in PCAT1. In addition, nine amino acids in the TMD were replaced by methionine in three different constructs.

Isolated peptidase domain (residues 1–148) was sub-cloned into vector pMCSG20 and expressed in RIL cells. The protein was purified by glutathione sepharose affinity chromatography, followed by on-column cleavage by TEV proteases.

The putative substrate, gene *Cthe_0535*, was cloned into vector pMCSG7 with an N-terminal 6× His tag and a C-terminal 3× Flag tag. Protein was expressed in RIL cells and purified on cobalt affinity resin (Clontech Laboratories). The His tag was removed by TEV protease and the protein was further purified by gel-filtration chromatography (Superdex 75 16/60) followed by anion-exchange chromatography (SOURCE 15Q).

The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Crystallization of wild-type PCAT1 in the absence of ATP. Preparation of a 35% bicelle stock was carried out as described previously^{22,34,35}. In brief, 0.26 g 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC; Avanti Polar Lipids) and 0.09 g 3-(cholamidopropyl)-dimethylammonio-2-hydroxyl-1-propanesulfonate (CHAPSO; Affymetrix) were mixed and dissolved in 600 µl of nanopure water. Multiple cycles of incubation on ice and at 42 °C with a few minutes of vortexing between cycles were performed to obtain a homogeneous bicelle stock. The final pH of the bicelle stock was adjusted to pH 7.0 with 1 M NaOH. Purified PCAT1 at 10 mg ml^{−1} was mixed with the pre-chilled bicelle stock at 4:1 (v/v) ratio on ice for 30 min, with additional 1.4 mM *N,N*-bis-(3-D-gluconamidopropyl) deoxycholate (deoxy Big CHAP). The final crystallization sample contains 8 mg ml^{−1} PCAT1, 2 mM UDM, 7% bicelles, and 1.4 mM deoxy Big CHAP.

Crystals in primitive orthorhombic space group *P*₂₁₂₁ were grown at 20 °C by vapour diffusion using a reservoir solution containing 100 mM sodium citrate, pH 5.2–5.4 and 19–22% PEG400. Well-formed crystals were transferred incrementally into a cryoprotectant containing 100 mM sodium citrate, pH 5.2, 20% PEG400, 10% glycerol, 7% bicelles, 2 mM UDM and 1.4 mM deoxy Big CHAP, and flash-frozen in liquid nitrogen. Crystals in C-centred orthorhombic space group *C*22₂ were grown in 100 mM sodium citrate, pH 5.0 and 10% PEG2000. Crystals were transferred incrementally into a cryoprotectant containing 100 mM sodium citrate, pH 5.0, 15% PEG2000, 30% glycerol, 7% bicelles, 2 mM UDM and 1.4 mM deoxy Big CHAP.

Crystallization of the E648Q mutant in complex with ATPγS. Protein sample at 15 mg ml^{−1} was mixed with 10 mM adenosine 5'-*o*-(3-thiotriphosphate) (ATPγS) and 7% bicelles for crystallization. Crystals were obtained at 20 °C by vapour diffusion using a reservoir solution containing 100 mM sodium citrate, pH 5.2–5.6, 7% PEG1500, and 20 mM CaCl₂.

Structural determination. X-ray diffraction data were collected at beamlines 23ID and 24ID of the Advanced Photon Source at 100 K. Data were processed with HKL2000 (ref. 36) and XDS³⁷ and scaled using Scalepack2MTZ in the CCP4

Program Suite^{38,39}. Ellipsoidal truncation and anisotropic scaling were performed by the Diffraction Anisotropy Server to improve data quality⁴⁰. The structure of wild-type PCAT1 was determined using molecular replacement in PHASER in the CCP4 Program Suite^{39,41}, using TM287 (PDB accession 3QF4, chain A) and the peptidase domain of *Streptococcus* ComA (PDB accession 3K8U) as search models. The model was built using Coot⁴². Refinement of the structure was performed in CNS 1.3 (ref. 43) and PHENIX⁴⁴, with translation-libration-screw (TLS)⁴⁵, two-fold non-crystallographic symmetry (NCS), and secondary structure restraints. In the ProCheck Ramachandran plot, 91.6% of residues were in favoured regions, and 8% of residues were in allowed regions. In the Molprobity Ramachandran plot, 95.5% of residues were in favoured regions, and 4.5% of residues were built with poor rotamers. Final refinement statistics are summarized in Extended Data Table 1.

The structure of the E648Q mutant was determined at 5.5 Å by molecular replacement in PHASER⁴¹ using the TMDs of the wild-type PCAT1 and the ATP-bound NBD dimer of ComA (PDB accession 3VX4) as search models. The model was built using COOT⁴² with correct sequence registry without side chains. Because of the low resolution, refinement was performed initially in CNS 1.3 (ref. 43) with the deformable elastic network (DEN) restraints⁴⁶, which uses the local structural information present in the reference model (the wild-type PCAT1 structure) and only change parts of the structure where diffraction data supported the change. The structure was further refined in PHENIX⁴⁴ with NCS, secondary structure, and reference model restraints. In the ProCheck Ramachandran plot, 92.2% of residues were in favoured regions, and 7.6% of residues were in allowed regions. In the Molprobity Ramachandran plot, 95.0% of residues were in favoured regions.

The electrostatic potential surface was calculated by the PDB2PQR server and PyMOL with the APBS plugin⁴⁷. All figures were generated with PyMOL⁴⁸.

Reconstitution of PCAT1 into proteoliposome. A lipid mixture composed of 3:1 (w/w) of 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (POPE; Avanti Polar Lipids) and 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho-(1'-rac-glycerol) (POPG; Avanti Polar Lipids) was used for reconstitution, using a previously reported protocol^{49–52}. Briefly, the dried lipid mixture was resuspended to 20 mg ml^{−1} in a reconstitution buffer containing 20 mM HEPES pH 7.5, 150 mM NaCl and 5 mM DTT, dispersed by sonication until solution was clear, and solubilized with a final concentration of 10 mM *n*-decyl-β-D-maltoside (DM). Detergent-solubilized PCAT1 at 2 mg ml^{−1} were mixed with the lipids at 1:10 (w/w) protein/lipid ratio and incubated at 22 °C for 2 h. Detergents were removed by dialysis against the reconstitution buffer at 4 °C for 5 days. The resulting proteoliposomes were flash-frozen with liquid nitrogen in 50 µl aliquots and stored at –80 °C.

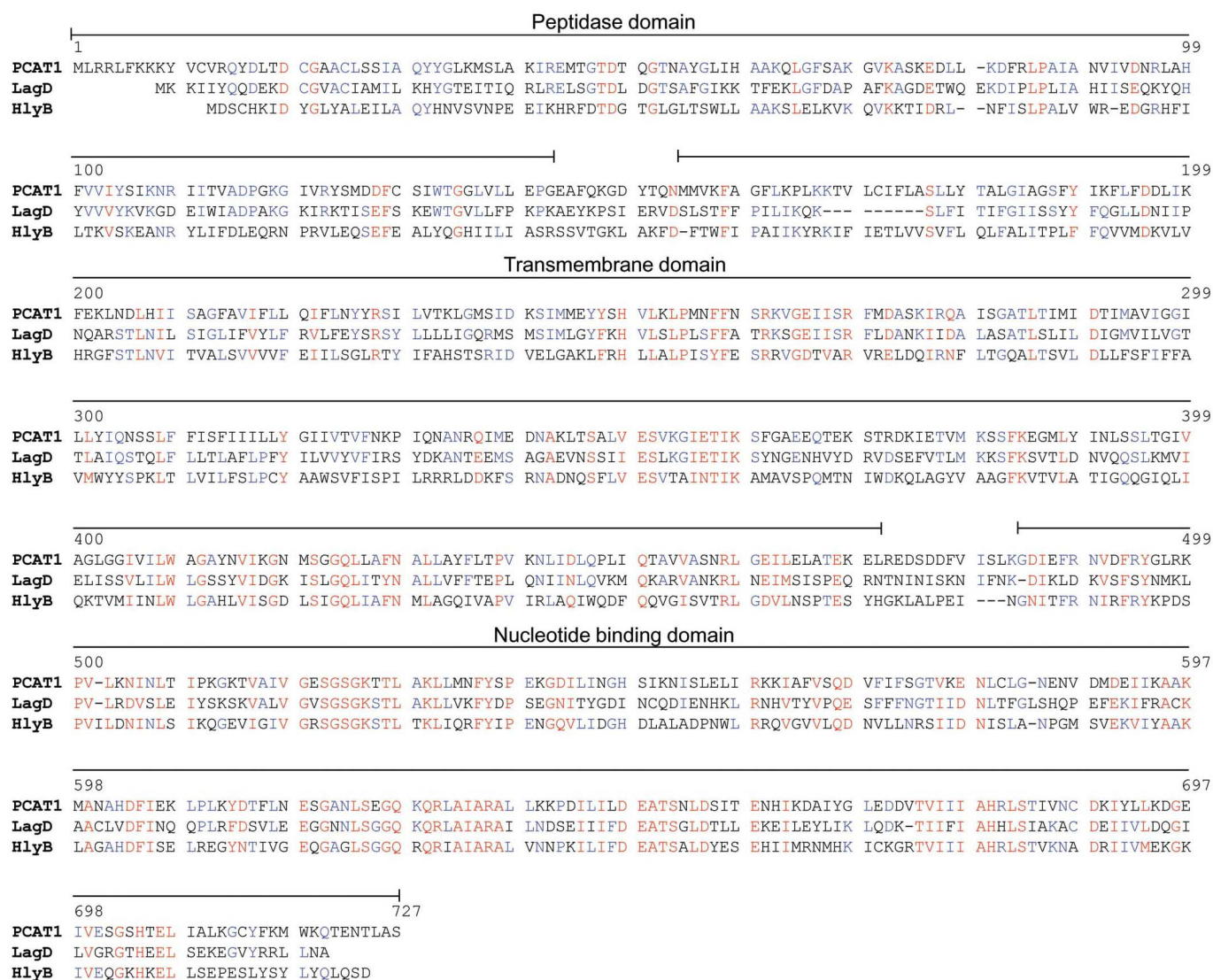
ATPase activities. The ATPase activity was determined using an ATP regeneration/NADH consumption-coupled system^{35,53,54}. Proteoliposomes containing 2.8 µg of PCAT1 were added into 30 µl of a reaction mixture containing 50 mM HEPES, pH 7.5, 60 µg ml^{−1} pyruvate kinase, 32 µg ml^{−1} lactate dehydrogenase, 4 mM phosphoenolpyruvate, 0.3 mM NADH and 1 mM MgCl₂. Different amounts of ATP/Mg²⁺ were added to initiate the reaction. The fluorescence of NADH was excited at 340 nm and recorded at 440 nm using the Synergy Neo HTS Multi-Mode Microplate Reader (BioTek). All measurements were performed at 37 °C or 50 °C for 30 min for at least three repeats. Calculations of the ATPase activities were based on the assumption that 50% of PCAT1 were incorporated into the proteoliposomes with the NBDs facing outside. Data were fitted to the Michaelis–Menten equation using Excel and the Solver Add-in.

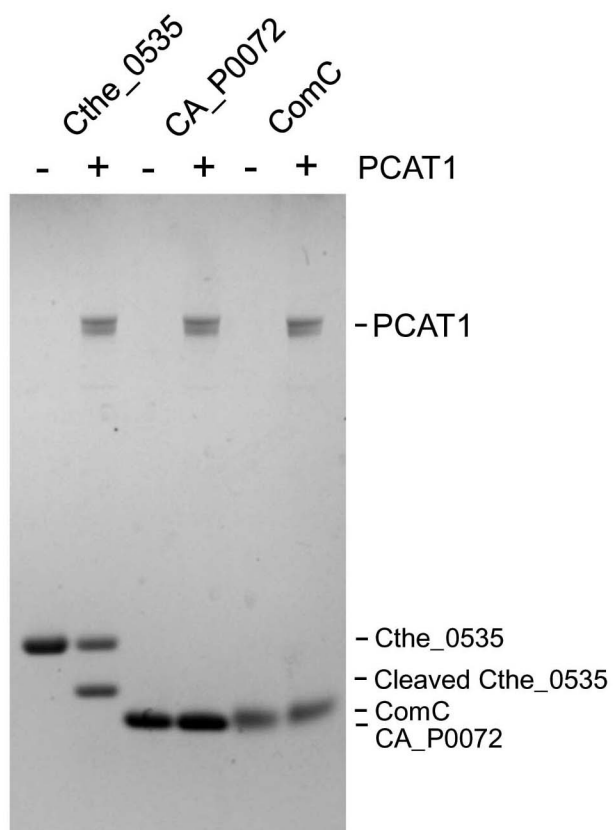
Peptidase cleavage assay. Proteoliposomes containing full-length PCAT1 or isolated peptidase domain (PEP; residues 1–148) were mixed with the substrate at 1:5 molar ratio in a reaction buffer containing 50 mM HEPES pH 7.0, 150 mM NaCl and 5 mM DTT. Cofactors were included as indicated at the following concentrations: 0.5 mM ATP, 1 mM MgCl₂, 10 mM EDTA or 5 mM ATPγS. Reactions were incubated at 37 °C for 5 h. Samples were analysed by NuPAGE 12% Bis-Tris Precast Gels in MES buffer (Life Technologies) and the intensities of protein bands were quantified using the ChemiDot-It² imaging system (UVP).

Pull-down assays. Substrate with a C-terminal 3×Flag tag was mixed with PEP, PCAT1_{151–727} or full-length PCAT1(C21A) mutant at 1:1 molar ratio in a reaction buffer containing 50 mM Tris, pH 7.0, 500 mM NaCl, 5 mM DTT and 2 mM UDM on ice for 20 min. Negative control samples were prepared by incubating PEP, PCAT1_{151–727} or PCAT1(C21A) without the substrate. Anti-Flag M2 Affinity Gel (Sigma-Aldrich) was then added to the samples and incubated for an additional 20 min. The M2 beads were washed extensively with the reaction buffer. GST-pull-down experiments using GST-tagged PEP and Glutathione Sepharose 4B resins were performed similarly. Cleaved substrate (residues

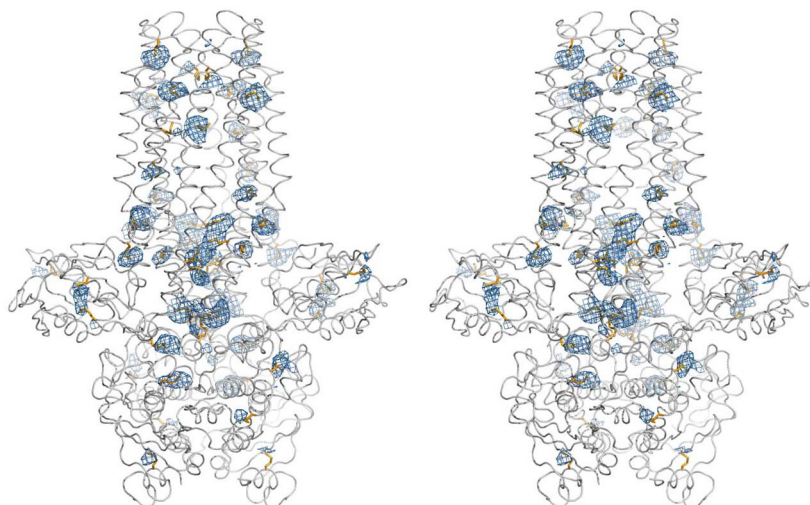
25–90) is a construct of the Cthe_0535 without the leader peptide. Leader–LBP is a construct in which the leader peptide is fused to the N terminus of a lanthanide-binding peptide. Both bound and unbound fractions were then analysed on SDS–PAGE.

34. Faham, S. & Bowie, J. U. Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. *J. Mol. Biol.* **316**, 1–6 (2002).
35. Chen, S., Oldham, M. L., Davidson, A. L. & Chen, J. Carbon catabolite repression of the maltose transporter revealed by X-ray crystallography. *Nature* **499**, 364–368 (2013).
36. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
37. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
38. French, G. S. & Wilson, K. S. On the treatment of negative intensity observations. *Acta Crystallogr. A* **34**, 517–525 (1978).
39. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
40. Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103**, 8060–8065 (2006).
41. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
42. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
43. Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
44. Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. A robust bulk-solvent correction and anisotropic scaling procedure. *Acta Crystallogr. D* **61**, 850–855 (2005).
45. Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D* **62**, 439–450 (2006).
46. Schröder, G. F., Levitt, M. & Brunger, A. T. Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464**, 1218–1222 (2010).
47. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
48. DeLano, W. L. in *The PyMOL User's Manual* (DeLano Scientific, 2002).
49. Heginbotham, L., LeMasurier, M., Kolmakova-Partensky, L. & Miller, C. Single *Streptomyces lividans* K⁺ channels: functional asymmetries and sidedness of proton activation. *J. Gen. Physiol.* **114**, 551–560 (1999).
50. Long, S. B., Tao, X., Campbell, E. B. & MacKinnon, R. Atomic structure of a voltage-dependent K⁺ channel in a lipid membrane-like environment. *Nature* **450**, 376–382 (2007).
51. Tao, X. & MacKinnon, R. Functional analysis of Kv1.2 and paddle chimera Kv channels in planar lipid bilayers. *J. Mol. Biol.* **382**, 24–33 (2008).
52. Brohawn, S. G., del Marmol, J. & MacKinnon, R. Crystal structure of the human K2P TRAAK, a lipid- and mechano-sensitive K⁺ ion channel. *Science* **335**, 436–441 (2012).
53. Scharschmidt, B. F., Keeffe, E. B., Blankenship, N. M. & Ockner, R. K. Validation of a recording spectrophotometric method for measurement of membrane-associated Mg- and NaK-ATPase activity. *J. Lab. Clin. Med.* **93**, 790–799 (1979).
54. Orelle, C., Ayvaz, T., Everly, R. M., Klug, C. S. & Davidson, A. L. Both maltose-binding protein and ATP are required for nucleotide-binding domain closure in the intact maltose ABC transporter. *Proc. Natl Acad. Sci. USA* **105**, 12837–12842 (2008).

Extended Data Figure 1 | Sequence alignment of PCAT1 from *Clostridium thermocellum*, LagD from *Lactococcus lactis*, and HlyB from *Escherichia coli*.

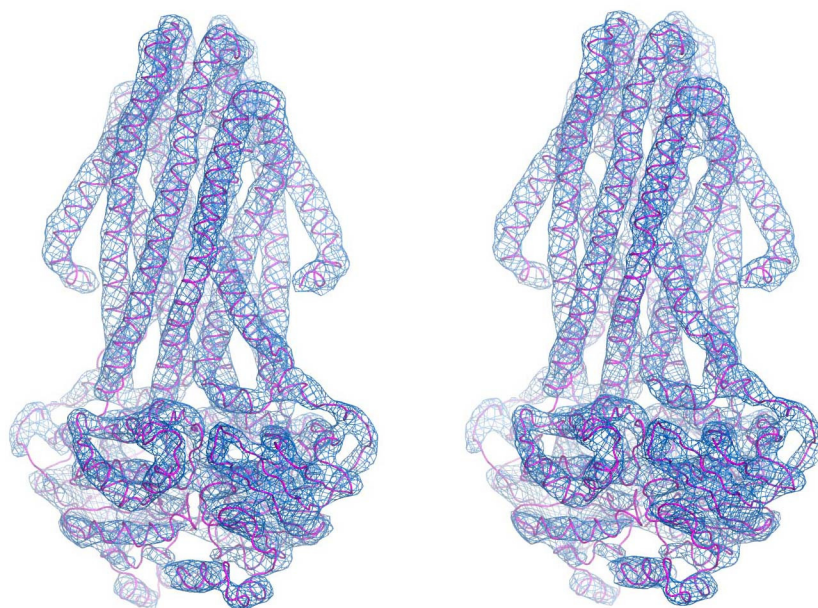


Extended Data Figure 2 | PCAT1 protease activities towards substrates of other Gram-positive bacteria. PCAT1 was able to cleave its putative substrate, Cthe_0535, from *C. thermocellum* at 37 °C for 2 h but showed no proteolytic activities towards CA_P0072 from *Clostridium acetobutylicum* or ComC from *Streptococcus pneumoniae*.

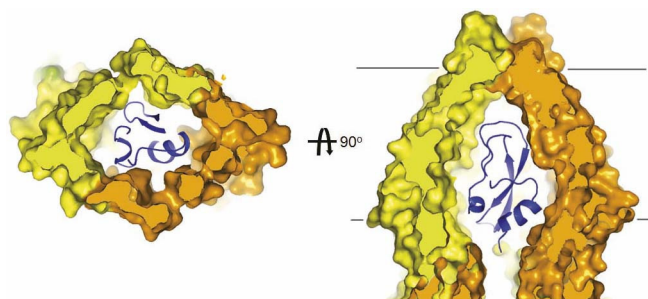


Extended Data Figure 3 | Anomalous difference Fourier electron density map. Stereoview of the backbone of SeMet-substituted PCAT1 (grey ribbon). Methionine residues are shown in orange sticks. The blue mesh contoured at 3σ represents the superimposed anomalous difference Fourier map calculated from data collected on four different PCAT1 constructs. A total of 28 selenium

sites were identified and used as markers to assist assignment of the sequence register. Out of the 21 native methionine residues, only two were not identified (Met 1 and Met 271), probably reflecting the conformational flexibility of these residues.



Extended Data Figure 4 | Stereoview of the final electron density map ($2F_o - F_o$, 1σ) of the E648Q mutant in complex with ATP γ S.



Extended Data Figure 5 | The TM tunnel in the ATP-free form is large enough to accommodate a small protein. The bovine pancreatic trypsin inhibitor (PDB accession 4PTI) is modelled into the TM tunnel of PCAT1, shown as a blue ribbon, to illustrate the size of the cavity.

Extended Data Table 1 | Data collection and refinement statistics (Molecular Replacement)

PCAT1	Wild-type		E648Q + ATP γ S
Data collection			
Space group	P2 ₁ 2 ₁ 2 ₁	C222 ₁	P4 ₂ 2 ₁ 2
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	87.59, 89.73, 296.59	138.38, 178.35, 90.21	230.00, 230.00, 89.4
α , β , γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	50.0-3.61 (3.74-3.61)	50.0-4.05 (4.19-4.05)	50-5.5 (5.7-5.5)
<i>R</i> _{sym}	0.093	0.089	0.083
<i>I</i> / σ <i>I</i>	16.0 (0.8)	17.41 (1.15)	31.1 (1.2)
Completeness (%)	96.0 (90.7)	94.2 (79.7)	99.2 (95.6)
Redundancy	7.1 (6.7)	5.0 (4.1)	4.9 (4.2)
Refinement			
Resolution (Å)	20.0-3.61 (3.77-3.61)		20.0-5.52 (6.90-5.52)
No. reflections	20686		6433
<i>R</i> _{work} / <i>R</i> _{free}	0.266/0.289 (0.343/0.343)		0.301/0.314 (0.429/0.436)
No. atoms			
Protein	9927		5574
ATP γ S	0		62
B-factors			
Protein	169.3		272.5
ATP γ S			321.7
R.m.s deviations			
Bond lengths (Å)	0.004		0.003
Bond angles (°)	0.079		0.068

Highest resolution shell is shown in parenthesis.

Antibody against early driver of neurodegeneration *cis* P-tau blocks brain injury and tauopathy

Asami Kondo^{1,2*}, Koorosh Shahpasand^{1,2*}, Rebekah Mannix³, Jianhua Qiu³, Juliet Moncaster⁴, Chun-Hau Chen^{1,2}, Yandan Yao^{1,2}, Yu-Min Lin^{1,2}, Jane A. Driver^{1,5}, Yan Sun⁶, Shuo Wei^{1,2}, Man-Li Luo^{1,2}, Onder Albayram^{1,2}, Pengyu Huang^{1,2}, Alexander Rotenberg⁶, Akihito Ryo⁷, Lee E. Goldstein⁴, Alvaro Pascual-Leone⁸, Ann C. McKee⁴, William Meehan⁹, Xiao Zhen Zhou^{1,2§} & Kun Ping Lu^{1,2§}

Traumatic brain injury (TBI), characterized by acute neurological dysfunction, is one of the best known environmental risk factors for chronic traumatic encephalopathy and Alzheimer's disease, the defining pathologic features of which include tauopathy made of phosphorylated tau protein (P-tau). However, tauopathy has not been detected in the early stages after TBI, and how TBI leads to tauopathy is unknown. Here we find robust *cis* P-tau pathology after TBI in humans and mice. After TBI in mice and stress *in vitro*, neurons acutely produce *cis* P-tau, which disrupts axonal microtubule networks and mitochondrial transport, spreads to other neurons, and leads to apoptosis. This process, which we term 'cistaosis', appears long before other tauopathy. Treating TBI mice with *cis* antibody blocks cistaosis, prevents tauopathy development and spread, and restores many TBI-related structural and functional sequelae. Thus, *cis* P-tau is a major early driver of disease after TBI and leads to tauopathy in chronic traumatic encephalopathy and Alzheimer's disease. The *cis* antibody may be further developed to detect and treat TBI, and prevent progressive neurodegeneration after injury.

Traumatic brain injury (TBI) is the leading cause of death and disability in children and young adults¹, and in the USA approximately 2.5 million people suffer TBI each year². Nearly 20% of the 2.3 million troops deployed by the military have sustained TBI³. Repetitive mild TBI (rmTBI), seen in contact sports, or even single moderate/severe TBI (ssTBI), seen in military blasts, may cause acute and potentially long-lasting neurological dysfunction, including the development of chronic traumatic encephalopathy (CTE)^{4–9}. TBI is also an established environmental risk factor for Alzheimer's disease^{7–12}. However, no treatment is currently available to prevent CTE or Alzheimer's disease.

CTE is characterized by neurofibrillary tangles made of hyperphosphorylated tau^{4–9}. Such tangles are also a hallmark of Alzheimer's disease and related neurodegenerative disorders, collectively termed tauopathies^{13,14}. Tauopathy spreads in brains^{15–19} and is reduced by immunotherapy against tauopathy epitopes^{20–22}. However, since little tauopathy is detectable acutely or subacutely after TBI in humans and mice^{5,7–9,23–25}, whether tauopathy is a cause or consequence of post-traumatic neurodegeneration is unknown.

We have identified a unique proline isomerase, Pin1, that inhibits tauopathy in Alzheimer's disease by converting the phosphorylated Thr231-Pro motif in tau (P-tau) from *cis* to *trans* in Alzheimer's disease cell and mouse models^{26–34}. In human Alzheimer's disease, Pin1 is inhibited by multiple mechanisms^{27,29,35–37}, whereas the Pin1 genetic polymorphism that prevents its downregulation is associated

with delaying Alzheimer's disease age of onset³⁸. In addition, Pin1 is located at a locus associated with late-onset Alzheimer's disease³⁹, P-tau appears early in pre-tangle Alzheimer's disease neurons⁴⁰, and its cerebrospinal fluid level correlates with memory loss in mild cognitive impairment and Alzheimer's disease⁴¹. We have developed antibodies that distinguish *cis* from *trans* P-tau and discovered that *trans* P-tau is physiological, promoting microtubule assembly, whereas the *cis* form is early pathogenic, leading to tauopathy in Alzheimer's disease⁴². Currently, it is unknown whether *cis* P-tau is present after TBI and if so, how to specifically eliminate it.

Robust *cis* P-tau in human CTE brains

We generated mouse monoclonal antibodies (mAbs) that, like our polyclonal antibodies⁴², were able to distinguish *cis* from *trans* tau. We identified a *cis* mAb clone, 113, and a *trans* mAb clone, 25, with no cross-reactivity (Extended Data Fig. 1a, b). Both clones reacted to a pT231-tau peptide, but not its non-phosphorylated counterpart (Extended Data Fig. 1a, b).

We determined antibody binding affinities using a Biacore assay. *Cis* and *trans* mAbs specifically recognized the P-tau peptide (Fig. 1a, b), with their binding constants (K_D) being 0.27 and 42.1 nM, respectively (Table 1). Their IgG isotypes were IgG2b and IgG1, respectively (Extended Data Fig. 1c). Immunofluorescence and immunoblotting analyses showed robust *cis* signals in the soma and neurites, and *trans* signals in the soma in tau-transgenic mice, but not in tau-null

¹Division of Translational Therapeutics, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA. ²Cancer Research Institute, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA. ³Division of Emergency Medicine, Children's Hospital Boston, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴Alzheimer's Disease Center, CTE Program, Boston University School of Medicine, Boston, Massachusetts 02118, USA. ⁵Geriatric Research Education and Clinical Center, VA Boston Healthcare System, Harvard Medical School, Boston, Massachusetts 02130, USA. ⁶Department of Neurology, Children's Hospital Boston, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁷Department of Microbiology, Yokohama City University School of Medicine, Yokohama 236-0004, Japan. ⁸Department of Neurology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA. ⁹Micheli Center for Sports Injury Prevention, Children's Hospital Boston, Harvard Medical School, Boston, Massachusetts 02115, USA.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

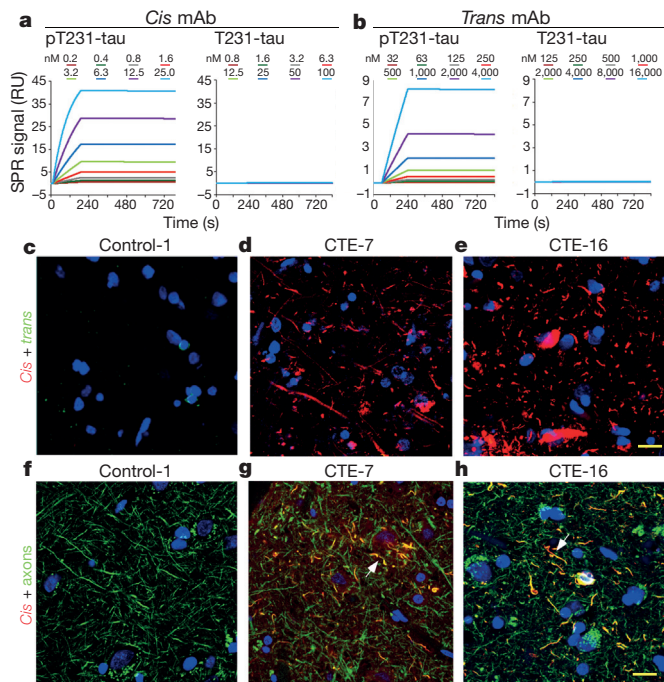


Figure 1 | Robust *cis*, but not *trans*, P-tau at diffuse axons in human CTE brains. **a, b,** *Cis* (**a**) or *trans* (**b**) mAb were immobilized on a sensor chip CM5 for surface plasmon resonance and their binding to pT231- or T231-tau peptide at different concentrations were recorded by SPR sensorgrams. **c–h,** The frontal cortex of neuropathologically verified human CTE brains and normal controls were subjected to double immunofluorescence with *cis* (red) or *trans* (green) mAbs (**c–e**), $n = 16$ for CTE and 8 for controls, or with *cis* pT231 (red) and the axonal marker tau (green), along with DNA dye (blue) (**f–h**), $n = 4$. Two typical *cis* P-tau immunostaining patterns are presented, with all cases being shown in Extended Data Fig. 1f, g. Arrows, colocalization; scale bars, 20 μm .

mice (Extended Data Fig. 1d, e). Thus, *cis* and *trans* mAb behave similarly to their polyclonal counterparts⁴².

Since the T231-tau phospho-epitope is identical among species, we performed double immunofluorescence with *cis* and *trans* mAbs on CTE brain tissues from 16 patients with a history of TBI exposure and 8 healthy controls⁶ (Supplementary Table 1). While *trans* mAb detected a few neurons in the soma in control and CTE brains, *cis* mAb detected no signal in control brains. However, robust *cis* mAb signals were observed in diffuse neurites in all CTE brains examined, with two typical patterns evident, distinguished by one with stronger *cis* P-tau signals, especially in soma (Fig. 1d, e and Extended Data Fig. 1f–h). *Cis* mAb co-localized with AT180 (recognizing pT231-tau), T22 (tau oligomers⁴³), AT8 (early tangles), and AT100 and Alz50 (mature tangles), but *trans* mAb did not co-localize with T22 (Extended Data Fig. 2a, b). *Cis* P-tau was more concentrated near blood vessels (Extended Data Fig. 2c), as expected⁶. *Cis* P-tau co-localized diffusely with the axonal marker tau, but not the dendritic marker MAP2, in CTE brains (Fig. 1g, h and Extended Data Fig. 2d). Thus, *cis* P-tau localizes primarily to diffuse axons in CTE brains.

Cis P-tau is the earliest TBI tau epitope

To determine the temporospatial characteristics of *cis* P-tau induction after TBI, we used TBI mouse models induced by impact²⁵ and blast⁵,

Table 1 | Binding affinities of *cis* and *trans* mAbs

mAb	Peptide	K_a (ms^{-1})	K_d (s^{-1})	K_D (nM)
<i>Cis</i>	pT231-tau	40,700	1.10×10^{-5}	0.27
<i>Cis</i>	T231-tau	0.17	1.00×10^{-5}	58,820
<i>Trans</i>	pT231-tau	250	1.05×10^{-5}	42.0
<i>Trans</i>	T231-tau	4.5	1.95×10^{-5}	4,333

The association rate constant (K_a), dissociation rate constant (K_d), and binding constant (K_D) of *cis* and *trans* mAbs towards pT231-tau or T231-tau peptide were determined by Biacore analysis.

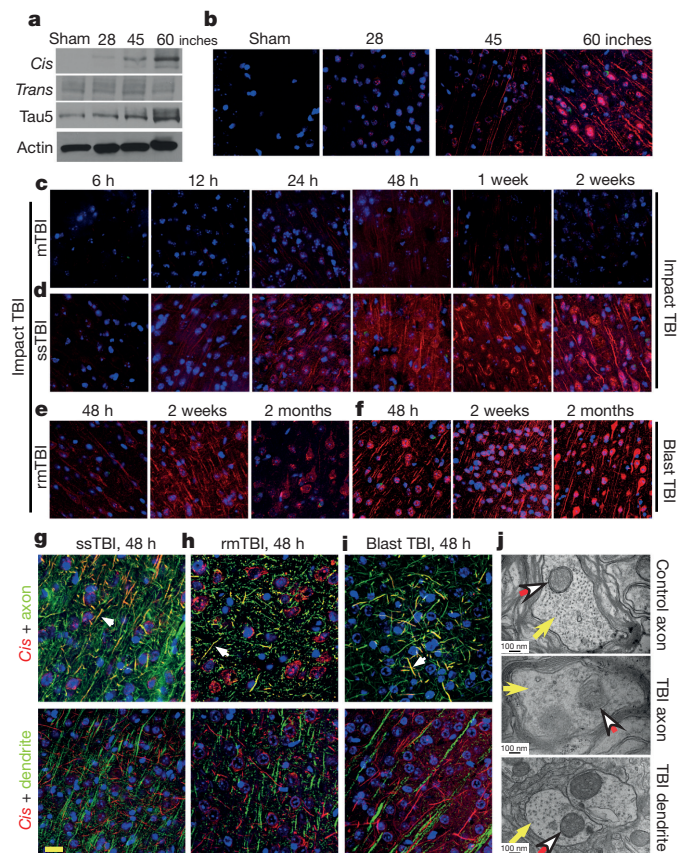


Figure 2 | While mTBI has moderate and transient effect, rmTBI, ssTBI or blast TBI leads to robust and persistent *cis* P-tau induction, notably in diffuse axons starting at 12–24 h. **a, b,** Mice were subjected to single TBI by a 54 g weight drop from varying heights, followed by immunoblotting (**a**) and immunofluorescence (**b**) to detect *cis* and *trans* P-tau 48 h later. *Cis*, red; *trans*, green; DNA, blue. **c–e,** Mice were subjected to single mTBI (**c**), ssTBI (**d**) or rmTBI (**e**), followed by immunofluorescence to detect *cis* and *trans* P-tau at different times after last injury. $n = 4$. **f,** Mice were subjected to blast-induced TBI, followed by immunofluorescence to detect *cis* and *trans* P-tau at different times. $n = 3$. **g–i,** ssTBI (**g**), rmTBI (**h**) and blast TBI (**i**) brain sections at 48 h after last injury were subjected to double immunofluorescence with *cis* pT231 (red) and axon marker neurofilament SMI312 (green) or dendrite marker MAP2 (green), along with DNA dye (blue). $n = 4$. Arrows, colocalization; magnification in **b–f**, $\times 63$; scale bars in **g–i**, 20 μm . **j,** ssTBI or sham mice were subjected to electron microscopy analysis 48 h after injury to examine the structure of microtubules (filled arrows) and mitochondria (open arrows) at axons and dendrites. Scale bars, 100 nm.

modelling sport- and military-related TBI, respectively. 48 h after impact TBI, *cis*, but not *trans*, P-tau was elevated in a severity-dependent manner, correlating with total tau (Fig. 2a, b and Extended Data Fig. 3a, b), and reflecting the stability of *cis* P-tau⁴². While single mild TBI (mTBI) moderately and transiently induced *cis* P-tau, which returned to the baseline by 2 weeks, ssTBI robustly and persistently induced *cis* P-tau, starting at 12 h and peaking at 48 h, but sustaining high levels over time (Fig. 2c, d and Extended Data Fig. 3c). Both rmTBI and blast TBI also induced robust and persistent *cis* P-tau induction, with more profound effects in the latter (Fig. 2e, f and Extended Data Fig. 3c).

Robust *cis* P-tau signals were detected 48 h after TBI without tau oligomers, aggregation or tangle epitopes (Extended Data Fig. 3d, e, and see results later). *Cis* P-tau localized mainly to axons, but not dendrites in impact and blast TBI models (Fig. 2g–i), as in CTE brains (Fig. 1g, h and Extended Data Fig. 2d). *Cis* P-tau expression was associated with axonal injury with marked disruption of microtubules and mitochondria, which was notably absent in dendrites (Fig. 2j),

consistent with the fact that TBI mainly affects axons⁴⁴. Thus, robust *cis* P-tau is induced acutely in axons after impact and blast TBI long before other forms of P-tau appear.

Cis P-tau spreads and is toxic after TBI

Further analysis showed *cis* P-tau spreading in the brain after TBI. *Cis* P-tau was mainly limited to the cortex from 24 h to 2 months after rmTBI, but 6 months after rmTBI, robust *cis* P-tau signals were detected in the cortex and other brain regions, including the hippocampus (Fig. 3a–c). Marked *cis* P-tau spread from the cortex to the hippocampus and even to the other side of the brain was observed after blast TBI (data not shown). To examine whether *cis* P-tau might

be neurotoxic, brain lysates prepared from rmTBI and sham mice 6 months post-injury were added to growing cultured neurons overnight. *Cis*, but not *trans*, P-tau was readily detected in neurons treated with rmTBI lysates, but not when treated with sham controls (Fig. 3d). Compared with untreated or sham-treated controls, neurons treated with rmTBI lysates had much higher rates of apoptosis, which was rescued by immunodepletion of total tau or *cis*, but not *trans*, mAb (Fig. 3e). Thus, after impact and blast TBI, *cis* P-tau is robustly induced, spreads through the brain over time, and induces apoptosis that is blocked by *cis* mAb.

Stress induces *cis* P-tau, blocked by mAb

To understand how *cis* P-tau induces apoptosis and spreads through the brain, we examined the *in vitro* response to neuronal stress induced by serum starvation or hypoxia. Both conditions induced *cis*, but not *trans*, P-tau (Fig. 4a and Extended Data Fig. 4a, d), well before tau aggregation (Extended Data Fig. 4e). The addition of

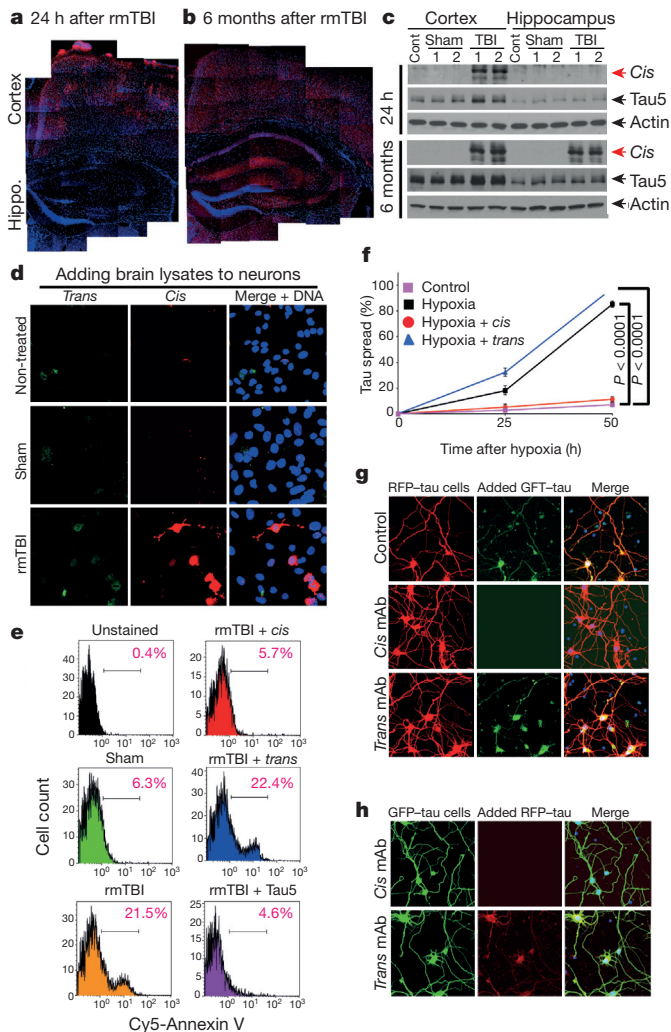


Figure 3 | *Cis* P-tau spreads in the brain after rmTBI, and spreads and causes neurotoxicity after neuronal stress *in vitro*, which are fully blocked by *cis*, but not *trans*, mAb. **a–c**, 24 h or 6 months after rmTBI, mouse brains were subjected to immunofluorescence (**a**, **b**) and immunoblotting (**c**) to detect *cis* P-tau in different brain regions. $n = 4$. **d**, **e**, Mouse brain lysates prepared from 6-month-post-rmTBI or sham controls were added to culture media of SY5Y neurons for 17 h directly or after immunodepletion with *cis* or *trans* mAb, followed by immunofluorescence with *cis* and *trans* mAbs or annexin V FACS. $n = 3$. **f**, SY5Y neurons stably expressing GFP-tau or RFP-tau were co-cultured and then treated with hypoxia or control in the presence or absence of *cis* or *trans* mAb for different times, followed by assaying cells expressing both GFP-tau and RFP-tau (mean \pm s.d.). P values, two-way ANOVA test. **g**, **h**, Primary mouse neurons were transfected with GFP-tau or RFP-tau, and then subjected to hypoxia treatment in the absence or presence of *cis* or *trans* mAb for 36 h. The resulting filtered soluble media from GFP-expressing neurons was added to RFP-expressing neurons (**g**) or vice versa (**h**), followed by detecting entry of added tau.

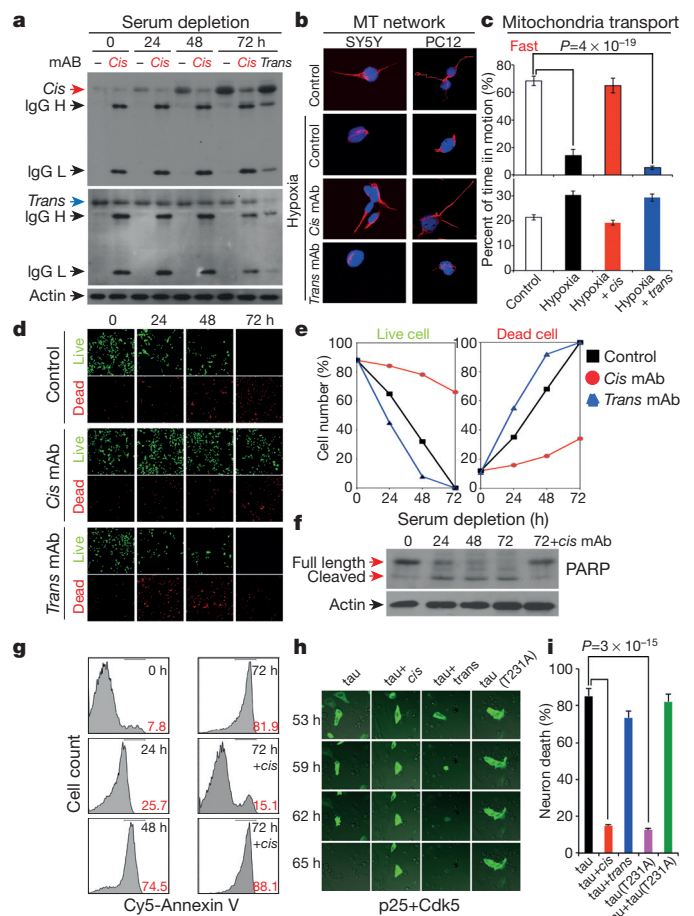


Figure 4 | Stressed neurons robustly produce *cis* P-tau leading to cistatosis, which is blocked by *cis* mAb, but enhanced by *trans* mAb. **a**, SY5Y cells were cultured without serum for different times in the absence and presence of *cis* or *trans* mAb, followed by immunoblotting for *cis* and *trans* P-tau. **b**, SY5Y and differentiated PC12 cells were treated with hypoxia in the absence and presence of *cis* or *trans* mAb for 48 h, followed by staining for microtubules. **c**, Differentiated PC12 cells were treated with hypoxia in the absence and presence of *cis* or *trans* mAb for 48 h, followed by live-cell microscopy to capture fast and slow transport of mitochondria along neurites. **d–g**, SY5Y cells were cultured without serum for different times in the absence and presence of *cis* or *trans* mAb, followed by live/dead cell assay (**d**, **e**) and apoptosis assays using PARP cleavage (**f**) and annexin V (**g**). **h**, **i**, SY5Y cells were co-transfected with GFP-tau or GFP-tau(T231A) and p25/Cdk5, followed by live-cell imaging to observe cell death of GFP-tau expressing cells over 65 h (**h**), with quantification being shown (**i**) (mean \pm s.d.). P values, ANOVA test.

stressed neuron lysates to neurons induced cell death, which was rescued by immunodepletion with *cis*, but not *trans*, mAb, as detected by the live/dead assay (Extended Data Fig. 4g). To examine whether *cis* P-tau is implicated in tau spreading, we generated SY5Y cells stably expressing green or red fluorescent protein-conjugated tau (GFP-tau or RFP-tau), and co-cultured them with or without stress. Without hypoxia, neurons continued to express either GFP-tau or RFP-tau, but rarely both proteins (Fig. 3f and Extended Data Fig. 5b). However, consistent with *cis* P-tau spreading in TBI brains (Fig. 3a–c and Extended Data Fig. 5a), hypoxia induced *cis* P-tau (Extended Data Fig. 4d) and caused progressive tau spreading, which was prevented by *cis*, but not *trans*, mAb (Fig. 3f and Extended Data Fig. 5b). Moreover, serum-starved neurons released *cis*, but not *trans*, P-tau at 40 h before neuronal death at 72 h (Extended Data Fig. 4h). Similar patterns of *cis* P-tau spread and neurotoxicity were also observed in primary neurons and blocked by *cis*, but not *trans*, mAb (Fig. 3g, h, Extended Data Fig. 5c, d). Thus, toxic *cis* P-tau is induced and spreads after neuronal stress, similar to TBI.

Cis mAb blocks cistauosis after stress

Given the ability of *cis* mAb to block tau from spreading and inducing apoptosis, we examined whether *cis* or *trans* mAb could affect intracellular P-tau after stress. Indeed, *cis* mAb entered neurons and effectively blocked time-dependent *cis* P-tau induction, without affecting *trans* following serum starvation (Fig. 4a and Extended Data Fig. 4a) or hypoxia (Extended Data Fig. 4d). Conversely, *trans* mAb reduced *trans*, but not *cis*, P-tau (Fig. 4a), indicating that the two isomers are not readily interchangeable, as suggested in TBI (Fig. 2) or CTE (Fig. 1), and Alzheimer's disease⁴².

Since Pin1 inhibition by downregulation^{27,29}, C113 oxidation³⁵ and S71 phosphorylation^{36,37} contributes to tauopathy in Alzheimer's disease, we asked whether such Pin1 inhibition contributes to *cis* P-tau induction after stress. *Cis* induction correlated highly with Pin1 downregulation after serum starvation and Pin1 C113 oxidation after hypoxia (Extended Data Fig. 6a, b). Pin1 S71 phosphorylation was also markedly elevated in TBI brains (Extended Data Fig. 6c). Moreover, Pin1 knockdown enhanced *cis* P-tau induction by hypoxia, which was eliminated by *cis* mAb (Extended Data Fig. 6d). Since Pin1 knockout induces P-tau accumulation only in old mice^{29,32,42} and stress activates Pro-directed kinases, increased tau phosphorylation may be also important for *cis* P-tau induction after stress or TBI.

Given the ability of *cis* mAb to ablate intracellular *cis* P-tau, we evaluated how *cis* mAb might enter neurons to remove *cis* P-tau. Tau mAbs enter neurons via Fcγ receptors⁴⁵ and mAbs trigger targeted protein degradation by the TRIM21-mediated proteasome pathway⁴⁶. Indeed, blocking Fcγ receptors prevented *cis* mAb from binding to or entering neurons (Extended Data Fig. 7a–c). Immunogold electron microscopy showed *cis* mAb on the outer cell surface and in intracellular vesicles (Extended Data Fig. 7d). TRIM21 knockdown⁴⁶, but not the autophagy inhibitor 3-methyladenine (3-MA), prevented *cis* mAb from ablating *cis* P-tau (Extended Data Fig. 7e–g). Thus, *cis* mAb likely enters neurons via Fcγ receptors to target *cis* P-tau degradation.

To examine the functional significance of neuronal *cis* P-tau induction and elimination, we investigated whether *cis* P-tau might affect the microtubule network and function since *cis*, but not *trans*, P-tau loses its microtubule function⁴². Hypoxia not only induced *cis* P-tau (Extended Data Fig. 4d), but also caused microtubule collapse in neurites, an effect that was rescued by *cis*, but not *trans*, mAb (Fig. 4b and Extended Data Fig. 4b). Measuring mitochondria movement along neurites showed that hypoxia stopped microtubule-based fast transport, but not actin-based slow movement (Fig. 4c and Supplementary Videos 1, 2). This mitochondria transport defect was restored by *cis* mAb, but not *trans* mAb (Fig. 4c and Supplementary Video 3), with the latter even causing neurite retraction (Supplementary Video 4), probably by *trans*-associated promotion of microtubule assembly⁴².

Serum starvation led to robust apoptosis by the time *cis* P-tau was highly induced, which was potentially rescued by *cis*, but not *trans* mAb, as detected by a live/dead assay (Fig. 4d, e), PARP cleavage (Fig. 4f and Extended Data Fig. 4c) and annexin V fluorescence activated cell sorting (FACS; Fig. 4g). Similar results were obtained with hypoxia (Extended Data Fig. 4f), even in primary neurons (Extended Data Fig. 5c, d). Thus, neuronal stress robustly induces *cis* P-tau, which disrupts axonal microtubules and organelle transport, spreads to other neurons and leads to apoptosis. These phenotypes, potentially rescued by *cis* but not *trans* mAb, are here termed 'cistauosis'.

To determine the importance of *cis* P-tau for neurotoxicity, we co-transfected neurons with GFP-tau or its T231A mutant and p25/Cdk5 phosphorylating tau on Thr231 and others^{27,47}. Co-expression of tau, but not its T231A mutant, with p25/Cdk5 increased *cis* P-tau, which was eliminated by *cis* mAb (Extended Data Fig. 8a). Importantly, most GFP-tau-, but not GFP-tau(T231A)-expressing cells were dead by 62 h, which was markedly blocked by *cis* mAb, but accelerated by *trans* mAb (Fig. 4h, i and Supplementary Videos 5, 6). Thus, *cis* P-tau is necessary and sufficient for P-tau to induce neurotoxicity.

Cis mAb potentially treats TBI and CTE

To evaluate the efficacy of *cis* mAb in treating TBI *in vivo*, we showed that *cis* or *trans* mAb were detected in brains 3 days after peripheral administration (Extended Data Fig. 9a). After treating ssTBI mice with *cis* mAb or IgG isotype control for 2 weeks, *cis* mAb effectively eliminated *cis* P-tau induction, both with and without pre-treatment (Fig. 5a and Extended Data Fig. 9b, d), and also potentially reversed post-TBI ultrastructural pathologies of axonal microtubules and mitochondria (Fig. 5c and Extended Data Fig. 9f), defective cortical axonal long-term potentiation (LTP) (Fig. 5d and Extended Data Fig. 9g), and even apoptosis (Extended Data Fig. 9c), which is observed after TBI even in humans⁴⁸.

To determine the impact of *cis* mAb on behavioural or functional outcomes after TBI, we treated ssTBI mice with *cis* mAb for 2 months. There was no difference in hippocampal-dependent spatial memory between IgG and *cis* mAb-treated TBI mice (Extended Data Fig. 9h). However, as *cis* P-tau was concentrated in the medial prefrontal cortex at this time point (Fig. 3a), we used the elevated plus maze, an innate anxiety/risk-taking paradigm that involves cortical circuitry⁴⁹ and is affected by TBI⁵⁰. All groups moved similar distances and times in the decision arm (Fig. 5e, f and Extended Data Fig. 10a–c). Sham mice stayed in the two closed or 'safe' arms (Fig. 5e, f and Extended Data Fig. 10a and Supplementary Video 7), but all IgG-treated ssTBI mice strikingly displayed 'risk-taking' behaviour, exploring the two open or 'aversive' arms (Fig. 5e, f and Extended Data Fig. 10a, b, and Supplementary Video 8), consistent with disinhibition likely due to a dysfunctional medial prefrontal cortex⁴⁹. By contrast, *cis* mAb-treated mice exhibited minimal risk-taking behaviour, similar to sham mice (Fig. 5e, f and Extended Data Fig. 10a, and Supplementary Video 9).

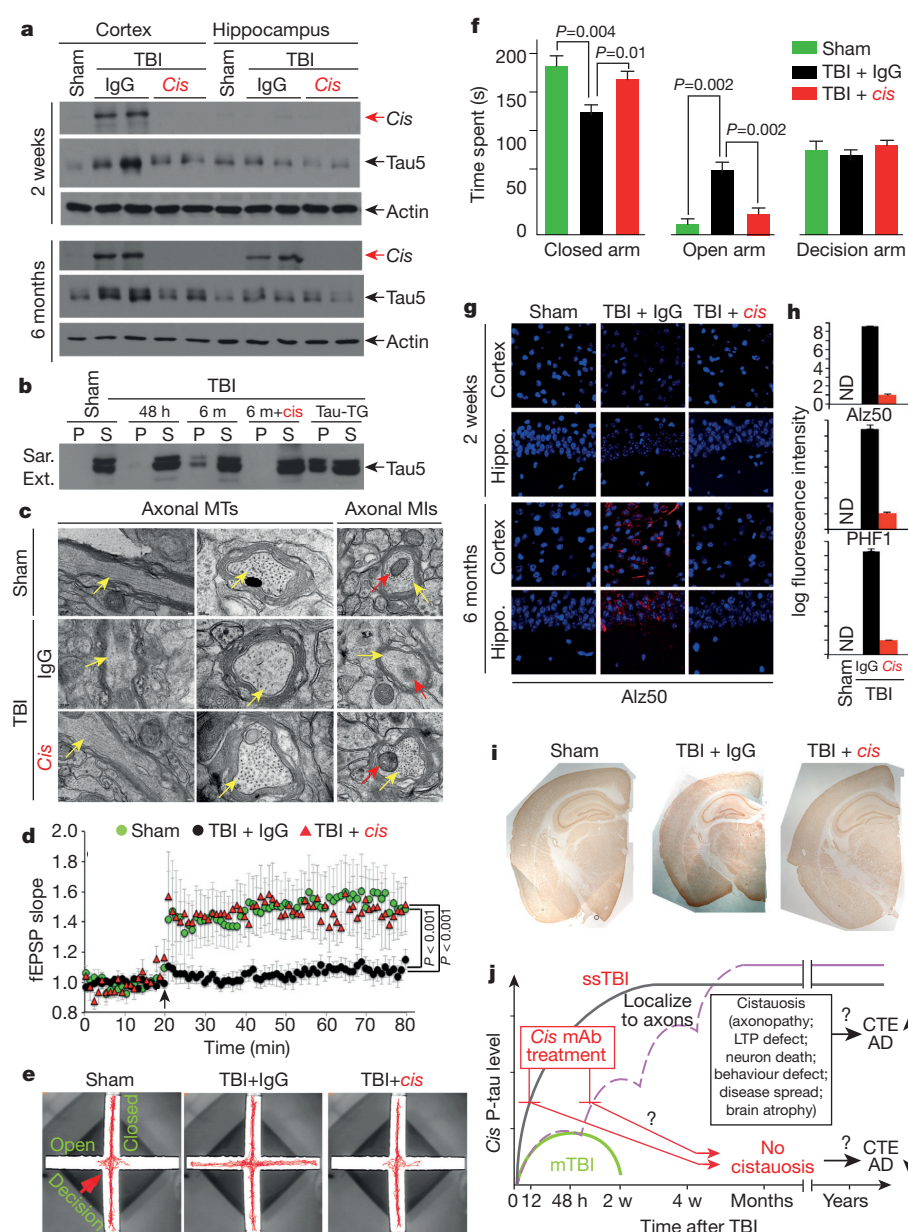
To examine the effects of *cis* mAb on tauopathy development and spread, and brain atrophy, hallmarks of CTE^{7–9}, we treated ssTBI mice with *cis* mAb for 6 months. *Cis* mAb effectively prevented tauopathy development and spread, as assayed by *cis* P-tau, tau oligomers, aggregation and tangle epitopes (Fig. 5a, b, g, h and Extended Data Figs 9d, e and 10d, e), and brain atrophy in the cortex and white matter (Fig. 5i and Extended Data Fig. 10f). Thus, *cis* mAb not only eliminates *cis* P-tau and cistauosis, but also prevents tauopathy development and spread, restores LTP and behavioural defects, and prevents brain atrophy after TBI (Fig. 5j).

Discussion

Here we used *cis* P-tau mAbs to demonstrate the presence of, and specifically eliminate, pathogenic *cis* P-tau in clinically relevant *in vitro* and *in vivo* models of sport- and military-related TBI. We detected robust *cis* P-tau signals after sport- and military-related TBI in humans and mice, and in stressed neurons. Following TBI or

Figure 5 | Treating ssTBI mice with *cis* mAb blocks early cistauosis, prevents tauopathy development and spread, and improves histopathological and functional outcomes.

a, b, ssTBI mice were treated with *cis* mAb or control IgG for times indicated, along sham mice as controls, followed by immunoblotting to detect *cis* P-tau (**a**), and sarcosyl extraction to detect tau aggregation (**b**). TG, transgenic; $n = 4$. **c, d,** After 2 week's treatment, ssTBI mice were subjected to electron microscopy for axonal structures of microtubules (yellow arrows) and mitochondria (red arrows) (c) or cortical fSPSP recording (**d**) (mean \pm s.e.m.). Black arrow, theta-burst application; $n = 15$ slices from 9 sham; $n = 9$ slices from 5 IgG or *cis* mAb mice. *P* values, one-way ANOVA with Bonferroni post hoc test. **e, f,** After 2 month's treatment, ssTBI mice were subjected to the elevated plus maze (**e**) and time spent in three arms was measured (**f**) (mean \pm s.e.m.). $n = 4$ for sham; $n = 7$ for IgG or *cis* mAb. *P* values, Student's *t*-test. **g, h,** After 2 week's or 6 month's treatment, ssTBI mice were subjected to immunofluorescence with tauopathy mAbs (**g**), with quantification in the hippocampus being shown (**h**) (means \pm s.d.). $n = 4$. **i,** After 6 month's treatment, ssTBI mice were immunostained with NeuN before determining brain thickness. $n = 4$. **j,** Unlike single mTBI, rmTBI or ssTBI causes robust and persistent *cis* P-tau induction within 12–24 h post-injury, which induces cistauosis, long before commonly known tauopathy and brain atrophy, hallmarks of CTE and Alzheimer's disease. *Cis* mAb not only blocks early cistauosis, but also prevents long-term neurodegeneration after TBI.



neuronal stress, *cis* P-tau induces cistauosis well before previously identified tauopathy is apparent. Treating TBI mice with *cis* mAb ablates *cis* P-tau and eliminates cistauosis, prevents the development of widespread tauopathy and restores histopathological and many functional outcomes of TBI. Cistauosis is an early precursor of previously described tauopathy and an early marker of neurodegeneration that can be blocked by *cis* mAb. We previously showed that *cis* P-tau has an early pathological role in Alzheimer's disease^{27–34,42}. Our current data provide a direct link from TBI to CTE and Alzheimer's disease, and suggest that cistauosis is a common early disease mechanism in TBI, CTE and Alzheimer's disease, and that *cis* P-tau and its mAb may be useful for early diagnosis, prevention and therapy for these devastating diseases (Fig. 5j).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 June 2014; accepted 11 June 2015.

Published online 15 July 2015.

1. Faul, M., Xu, L., Wald, M. M. & Coronado, V. G. Traumatic brain injury in the United States: emergency department visits, hospitalizations, and deaths, 2002–2006.

http://www.cdc.gov/traumaticbraininjury/tbi_ed.html (Centers for Disease Control and Prevention, 2010).

- Centers for Disease Control and Prevention. CDC grand rounds: reducing severe traumatic brain injury in the United States. *MMWR Morb. Mortal. Wkly Rep.* **62**, 549–552 (2013).
- Tanielian, T. *et al.* Invisible wounds of war: psychological and cognitive injuries, their consequences, and services to assist recovery. <http://www.rand.org/pubs/monographs/MG720> (RAND Corporation, 2008).
- Omalu, B. I. *et al.* Chronic traumatic encephalopathy in a National Football League player. *Neurosurgery* **57**, 128–134 (2005).
- Goldstein, L. E. *et al.* Chronic traumatic encephalopathy in blast-exposed military veterans and a blast neurotrauma mouse model. *Sci. Transl. Med.* **4**, 134ra160 (2012).
- McKee, A. C. *et al.* The spectrum of disease in chronic traumatic encephalopathy. *Brain* **136**, 43–64 (2013).
- Smith, D. H., Johnson, V. E. & Stewart, W. Chronic neuropathologies of single and repetitive TBI: substrates of dementia? *Nature Rev. Neurol.* **9**, 211–221 (2013).
- DeKosky, S. T., Blennow, K., Ikonomic, M. D. & Gandy, S. Acute and chronic traumatic encephalopathies: pathogenesis and biomarkers. *Nature Rev. Neurol.* **9**, 192–200 (2013).
- Blennow, K., Hardy, J. & Zetterberg, H. The neuropathology and neurobiology of traumatic brain injury. *Neuron* **76**, 886–899 (2012).
- Mortimer, J. A. *et al.* Head trauma as a risk factor for Alzheimer's disease: a collaborative re-analysis of case-control studies. EURODEM Risk Factors Research Group. *Int. J. Epidemiol.* **20** (Suppl. 2), S28–S35 (1991).
- Guo, Z. *et al.* Head injury and the risk of AD in the MIRAGE study. *Neurology* **54**, 1316–1323 (2000).

12. Nordström, P., Michaelsson, K., Gustafson, Y. & Nordström, A. Traumatic brain injury and young onset dementia: a nationwide cohort study. *Ann. Neurol.* **75**, 374–381 (2014).
13. Ballatore, C., Lee, V. M. & Trojanowski, J. Q. Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. *Nature Rev. Neurosci.* **8**, 663–672 (2007).
14. Mandelkow, E. M. & Mandelkow, E. Biochemistry and cell biology of tau protein in neurofibrillary degeneration. *Cold Spring Harb. Perspect. Med.* **2**, a006247 (2012).
15. Clavaguera, F. *et al.* Transmission and spreading of tauopathy in transgenic mouse brain. *Nature Cell Biol.* **11**, 909–913 (2009).
16. de Calignon, A. *et al.* Propagation of tau pathology in a model of early Alzheimer's disease. *Neuron* **73**, 685–697 (2012).
17. Liu, L. *et al.* Trans-synaptic spread of tau pathology *in vivo*. *PLoS ONE* **7**, e31302 (2012).
18. Clavaguera, F. *et al.* Peripheral administration of tau aggregates triggers intracerebral tauopathy in transgenic mice. *Acta Neuropathol.* **127**, 299–301 (2014).
19. Clavaguera, F., Hench, J., Goedert, M. & Tolnay, M. Invited review: prion-like transmission and spreading of tau pathology. *Neuropathol. Appl. Neurobiol.* **41**, 47–58 (2015).
20. Asuni, A. A., Boutajangout, A., Quartermain, D. & Sigurdsson, E. M. Immunotherapy targeting pathological tau conformers in a transgenic mouse model reduces brain pathology with associated functional improvements. *J. Neurosci.* **27**, 9115–9129 (2007).
21. Rosenmann, H. Immunotherapy for targeting tau pathology in Alzheimer's disease and tauopathies. *Curr. Alzheimer Res.* **10**, 217–228 (2013).
22. Sigurdsson, E. M. Tau immunotherapy and imaging. *Neurodegener. Dis.* **13**, 103–106 (2014).
23. Smith, C., Graham, D. I., Murray, L. S. & Nicoll, J. A. Tau immunohistochemistry in acute brain injury. *Neuropathol. Appl. Neurobiol.* **29**, 496–502 (2003).
24. Johnson, V. E., Stewart, W. & Smith, D. H. Widespread tau and amyloid-beta pathology many years after a single traumatic brain injury in humans. *Brain Pathol.* **22**, 142–149 (2012).
25. Mannix, R. *et al.* Clinical correlates in an experimental model of repetitive mild brain injury. *Ann. Neurol.* **74**, 65–75 (2013).
26. Lu, K. P., Hanes, S. D. & Hunter, T. A human peptidyl-prolyl isomerase essential for regulation of mitosis. *Nature* **380**, 544–547 (1996).
27. Lu, P. J., Wulf, G., Zhou, X. Z., Davies, P. & Lu, K. P. The prolyl isomerase Pin1 restores the function of Alzheimer-associated phosphorylated tau protein. *Nature* **399**, 784–788 (1999).
28. Zhou, X. Z. *et al.* Pin1-dependent prolyl isomerization regulates dephosphorylation of Cdc25C and tau proteins. *Mol. Cell* **6**, 873–883 (2000).
29. Liou, Y.-C. *et al.* Role of the prolyl isomerase Pin1 in protecting against age-dependent neurodegeneration. *Nature* **424**, 556–561 (2003).
30. Pastorino, L. *et al.* The prolyl isomerase Pin1 regulates amyloid precursor protein processing and amyloid-beta production. *Nature* **440**, 528–534 (2006).
31. Lu, K. P. & Zhou, X. Z. The prolyl isomerase PIN1: a pivotal new twist in phosphorylation signalling and human disease. *Nature Rev. Mol. Cell Biol.* **8**, 904–916 (2007).
32. Lim, J. *et al.* Pin1 has opposite effects on wild-type and P301L tau stability and tauopathy. *J. Clin. Invest.* **118**, 1877–1889 (2008).
33. Lee, T. H., Pastorino, L. & Lu, K. P. Peptidyl-prolyl cis-trans isomerase Pin1 in aging, cancer and Alzheimer's disease. *Expert Rev. Mol. Med.* **13**, e21 (2011).
34. Driver, J. A., Zhou, X. Z. & Lu, K. P. Pin1 dysregulation helps to explain the inverse association between cancer and Alzheimer's disease. *Biochim. Biophys. Acta*. <http://dx.doi.org/10.1016/j.bbagen.2014.12.025> (2015).
35. Chen, C. H. *et al.* Pin1 cysteine-113 oxidation inhibits its catalytic activity and cellular function in Alzheimer's disease. *Neurobiol. Dis.* **76**, 13–23 (2015).
36. Lee, T. H. *et al.* Death associated protein kinase 1 phosphorylates Pin1 and inhibits its prolyl isomerase activity and cellular function. *Mol. Cell* **42**, 147–159 (2011).
37. Kim, B. M. *et al.* Death-associated protein kinase 1 has a critical role in aberrant tau protein regulation and function. *Cell Death Dis.* **5**, e1237 (2014).
38. Ma, S. L. *et al.* A PIN1 polymorphism that prevents its suppression by AP4 associates with delayed onset of Alzheimer's disease. *Neurobiol. Aging* **33**, 804–813 (2012).
39. Wijsman, E. M. *et al.* Evidence for a novel late-onset Alzheimer disease locus on chromosome 19p13.2. *Am. J. Hum. Genet.* **75**, 398–409 (2004).
40. Luna-Muñoz, J., Chavez-Macias, L., Garcia-Sierra, F. & Mena, R. Earliest stages of tau conformational changes are related to the appearance of a sequence of specific phospho-dependent tau epitopes in Alzheimer's disease. *J. Alzheimers Dis.* **12**, 365–375 (2007).
41. Hampel, H. *et al.* Total and phosphorylated tau protein as biological markers of Alzheimer's disease. *Exp. Gerontol.* **45**, 30–40 (2010).
42. Nakamura, K. *et al.* Proline isomer-specific antibodies reveal the early pathogenic tau conformation in Alzheimer's disease. *Cell* **149**, 232–244 (2012).
43. Lasagna-Reeves, C. A. *et al.* Identification of oligomers at early stages of tau aggregation in Alzheimer's disease. *FASEB J.* **26**, 1946–1959 (2012).
44. Johnson, V. E., Stewart, W. & Smith, D. H. Axonal pathology in traumatic brain injury. *Exp. Neurol.* **246**, 35–43 (2013).
45. Congdon, E. E., Gu, J., Sait, H. B. & Sigurdsson, E. M. Antibody uptake into neurons occurs primarily via clathrin-dependent Fcγ receptor endocytosis and is a prerequisite for acute tau protein clearance. *J. Biol. Chem.* **288**, 35452–35465 (2013).
46. Mallery, D. L. *et al.* Antibodies mediate intracellular immunity through tripartite motif-containing 21 (TRIM21). *Proc. Natl Acad. Sci. USA* **107**, 19985–19990 (2010).
47. Cruz, J. C., Tseng, H. C., Goldman, J. A., Shih, H. & Tsai, L. H. Aberrant Cdk5 activation by p25 triggers pathological events leading to neurodegeneration and neurofibrillary tangles. *Neuron* **40**, 471–483 (2003).
48. Williams, S. *et al.* In situ DNA fragmentation occurs in white matter up to 12 months after head injury in man. *Acta Neuropathol.* **102**, 581–590 (2001).
49. Adhikari, A., Topiwala, M. A. & Gordon, J. A. Single units in the medial prefrontal cortex with anxiety-related firing patterns are preferentially influenced by ventral hippocampal activity. *Neuron* **71**, 898–910 (2011).
50. Schwarzbald, M. L. *et al.* Effects of traumatic brain injury of different severities on emotional, cognitive, and oxidative stress-related parameters in mice. *J. Neurotrauma* **27**, 1883–1893 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank T. Hunter and M. Zeidel for advice; S. Hagen for Microscopy Facility (NIH grant S10 RR017927) and P. Davies for tauopathy antibodies. C.-H.C., Y.-M.L., J.A.D. and S.W. are recipients of NIA-funded T32 Translational Research in Aging Training Grant, National Science Council Postdoctoral Fellowship from Taiwan, a VA Career Development Award, and Susan G. Komen postdoctoral fellowship, respectively. R.M. is supported by Boston Children's Hospital Pilot Grant Award and NIH training grant T32HD040128, and A.P.-L. and W.M. by NFLPA. The CTE and blast samples used are supported by grants from NIH (U01NS086659-01, P30AG13846), VA, Sports Legacy Institute, Andlinger Foundation, NFL and WWE. The work is supported by NIH grants R01AG029385, R01CA167677, R01HL111430 and R01AG046319, and Alzheimer's Association grant DVT-14-322623 to K.P.L. and BIDMC and NFLPA pilot grants to K.P.L. and X.Z.Z.

Author Contributions A.K. and K.S. designed the studies, performed the experiments, and wrote the manuscript; R.M. helped design and conduct experiments and analysed the data on impact TBI mouse models and wrote the manuscript; J.Q. and W.M. helped with impact TBI experiments, J.M. and L.E.G. helped with blast TBI experiments and edited the manuscript; A.C.M. provided human brains and edited the manuscript; Y.S. and A.Ro. performed field excitatory postsynaptic potential (fEPSP) recording; C.-H.C., Y.Y., Y.-M.L., J.A.D., S.W., M.-L.L., O.A. and P.H. provided technical assistance; A.Ry. provided assistance for developing mAbs; A.P.-L. advised the project; X.Z.Z. originally discovered the procedures for generating *cis* and *trans* antibodies; and X.Z.Z. and K.P.L. conceived and supervised the project, designed the studies, analysed the data, and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.P.L. (klu@bidmc.harvard.edu) or X.Z.Z. (xzhou@bidmc.harvard.edu).

METHODS

Mouse mAb production. *Cis* and *trans* mouse mAbs were produced using the general strategy that we used to generate polyclonal *cis* and *trans* antibodies, as described⁴². Briefly, Balb/c female mice (2–3 months old) obtained from the Jackson Laboratories (Bar Harbour, ME) were immunized with 100 µg of pThr231-Homoproline (pThr231-Pip) tau peptide (CKKVAVVRpT(Pip)PKSPSSAK) that was coupled to KLH with N-terminal Cys mixed with complete Freund's adjuvant and boosted twice. The titration of antibody production was monitored using ELISA. When sufficient titration of antibody was produced, splenocytes were isolated and fused with SP2/0 myeloma cells to produce hybridoma cell lines, followed by screening for positive clones using ELISA for *cis* and *trans* mAbs. When positive clones were identified, they were subcloned by a limited dilution to generate single pure clones. mAbs were produced by injecting 2 ml of 2.5×10^6 cells per ml hybridoma cells into nude mice intraperitoneally to collect ascites, followed by purifying mAbs from ascites using antibody purification kit (Pierce) and their specificity were fully characterized. All these animal experiments were approved by Beth Israel Deaconess Medical Center IACUC and complied with the NIH Guide for the Care and Use of Laboratory Animals.

ELISA assays. ELISA assays were performed using wild-type phosphorylated Thr231-Pro tau (KVAVVRpTPPKSPS), non-phosphorylated Thr231 tau (KVAVVRTPPKSPS), *cis* locked phosphorylated Thr231-Dmp tau (KVAVVRpT(5,5-dimethyl-L-proline)PKSPS) and *trans* lock phosphorylated Thr231-Ala tau (P232A)(KVAVVRpTAPKSPS) peptides, as described⁴². Briefly, peptides at various concentrations in 2,2,2-trifluoroethanol (50 µl) were plated onto maxi soap ELISA plate and dried up at 37 °C overnight. After blocking with buffer containing 5% milk, 0.4% bovine serum albumin and 0.05% Tween 20 in Tris-buffered saline, the *cis* or *trans* mAbs at various dilutions in 5% milk, 0.4% bovine serum albumin and 0.05% Tween 20 in Tris-buffered saline (50 µl) was loaded and incubated at room temperature for 2 h, followed by incubation with horseradish peroxidase (HRP)-conjugated anti-rabbit IgG in 5% milk, 0.4% bovine serum albumin and 0.05% Tween 20 in Tris-buffered saline (50 µl) for 1 h. The ELISA plates were washed 4 times with buffer containing 0.4% BSA and 0.05% Tween 20 in TBS after each step. The signals were detected by incubating with TMB substrate solution and were measured by Wallac 1420 software at 450 nm.

Surface plasmon resonance. Surface plasmon resonance experiments were performed on a BIAcore 3000 surface plasmon resonance instrument (GE Healthcare-BIAcore) as described by the manufacturer. Briefly, BIAcore sensor chip CM-5 was activated by using EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide) and NHS (*N*-hydroxysuccinimide) in a 1:1 ratio for 7 min. Anti-mouse IgG (Fc) (GE healthcare) was immobilized at pH 5 on flow cells 1 and 2, followed by the capture of $3.7 \mu\text{g ml}^{-1}$ of *cis* or *trans* mAb in 10 mM sodium acetate with a flow rate of $5 \mu\text{l min}^{-1}$. Then all tau peptides were injected at different concentrations in filtered, degassed 0.01 M HEPES buffer, 0.15 M NaCl, 0.005% surfactant P20, pH 7.4 at a flow rate of $50 \mu\text{l min}^{-1}$ for 3 min on both flow cells 1 and 2 and allowed to dissociate for 10 min. All samples were run in duplicate. After each run with a single antibody concentration, the surface was totally regenerated by 10 mM glycine pH 1.7 flow rate $10 \mu\text{l min}^{-1}$ for 5 s. Data analysis was performed by using BIAevaluation software (GE healthcare-BIAcore).

Immunoblotting analysis and immunodepletion experiment. Immunoblotting analysis was carried out as described⁴². Briefly, brain tissues or culture cells were lysed in RIPA buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 2 mM EDTA, 1% NP 40, 0.1% SDS, 0.5% Na-deoxycholate, 50 mM NaF) containing proteinase inhibitors and then mixed with the SDS sample buffer and loaded onto a gel after boiling. The proteins were resolved by polyacrylamide gel electrophoresis and transferred to PVDF membrane. After blocking with 5% milk in TBST (10 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.1% Tween 20) for 1 h, the membrane was incubated with primary antibodies (*cis* and *trans* mAbs), Tau5 (Biosource Camarillo, CA), α -tubulin (Sigma, St. Louis, MO) and β -actin antibodies (Sigma, St. Louis, MO) in 5% milk in TBST overnight at 4 °C. Then, the membranes were incubated with HRP-conjugated secondary antibodies in 5% milk in TBST. The signals were detected using chemiluminescence reagent (Perkin Elmer, San Jose, CA). The membranes were washed 4 times with TBST after each step. To deplete *cis* or *trans* P-tau from lysates, brain or cell lysates were mixed with the *cis* or *trans* mAb antibody at $425 \mu\text{g ml}^{-1}$ in RIPA buffer containing proteinase inhibitors for 3 h at 4 °C and then mixed with protein A/G Sepharose for 1 h at 4 °C, followed by collecting the supernatants for experiments. The supernatants were dialysed against phosphate buffer saline (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 , 1.8 mM KH_2PO_4) overnight before cell culture application. Immunoblotting results were quantified using Quantity One from BioRad.

Sarkosyl extraction. Isolation of sarkosyl-insoluble and soluble fractions of cells and brain tissues was performed as described^{29,32,42}, with slight modifications. Briefly, whole brains of mice were homogenized by polytron in 10 volumes of

buffer H (10 mM Tris-HCl (pH 7.5) containing 0.8 M NaCl, 1 mM EGTA, and 1 mM dithiothreitol). The cell extraction was also performed with a convenient amount of buffer H (200 µl per 35 mm culture dish) and sonication. The samples were spun at 100,000g for 30 min at 4 °C. Another 2 ml of buffer H was added to the pellet and the samples were homogenized again by polytron, incubated in 1% Triton X-100 at 37 °C for 30 min. Following the incubation, the samples were spun at 100,000g for 30 min at 4 °C, the pellet was homogenized by polytron in 1 ml of buffer H and was then incubated in 1% sarkosyl at 37 °C for 30 min and spun at 100,000g for 30 min at 4 °C. The supernatant was then collected (sarkosyl-soluble fraction). Detergent-insoluble pellets were extracted in 100 µl of urea buffer (8 M urea, 50 mM Tris-HCl (pH 7.5)), sonicated, and spun at 100,000g for 30 min at 4 °C. The supernatant was then collected (sarkosyl-insoluble fraction). The protein concentrations of extracts were determined by BCA assay (Thermo Scientific). Sarkosyl-insoluble and -soluble fractions were run on SDS-PAGE gels.

Immunostaining analysis. The primary antibodies used were *cis* and *trans* mAb, tau tangle-related mAbs AT180, AT8, AT100 (all from Innogenetics, Alpharetta, GA), oligomeric tau T22 polyclonal antibodies (EMD Millipore, Billerica, MA), PHF1 and Alz50 (gifts from P. Davies), anti-tau rabbit mAb (E178, Abcam) and anti-neurofilament mouse mAb (SMI-312, IgG1, Abcam) for labelling axons, and anti-MAP2 mAb (SMI-52, IgG1, Abcam) for labelling dendrites. Immunofluorescence staining of mouse and human brains was done essentially as described^{29,32,42}. After treatment with 0.3% hydrogen peroxide, slides were briefly boiled in 10 mM sodium citrate, pH 6.0, for antigen enhancement. The sections were incubated with primary antibodies overnight at 4 °C. Then, biotin-conjugated secondary antibodies (Jackson ImmunoResearch), streptavidin-conjugated HRP (Invitrogen) were used to enhance the signals. For double immunofluorescence staining, the sections were also incubated with Alexa Fluor 488 or 568 conjugated isotype-specific secondary antibodies (Jackson ImmunoResearch, West Grove, PA) for 1 h at room temperature. Manufacturer-supplied blocking buffer (Invitrogen) was used for each reaction. The sections were washed 4 times with TBS after each step. Labelled sections were visualized with a Zeiss confocal microscope. The gain of confocal laser was set at the level where there are no fluorescence signals including autofluorescence in sections without primary antibody but with secondary antibody. Immunostaining images and their colocalization were quantified using Volocity 6.3 from Perkin Elmer and Fiji/ImageJ Coloc 2, respectively.

Electronic microscopy. Sham and TBI mouse models treated with either control IgG or *cis* mAb were perfused with a fixative solution, a mixture of 15% picric acid (13% saturated solution; Sigma, St. Louis, MO, USA), 4% paraformaldehyde (Electron Microscopy Sciences, Hatfield, PA, USA), and 0.1% glutaraldehyde (electron microscopy grade 50% solution; Electron Microscopy Sciences) dissolved in PEM buffer (0.1 M PIPES, pH 7.2, 1 mM EGTA, 1 mM MgCl_2). Perfused brains were removed, sliced and kept in the same fixative for further 4 h at 4 °C. The samples were processed for electron microscopic observation as described^{51,52}. Specimens were examined with a JEM-1010 transmission electron microscope (JEOL). For immunogold staining, SY5Y cells were treated with *cis* mAb for 18 h, trypsinized and collected by centrifugation, followed by fixation with 4% PFA. Samples were dehydrated with ethanol, processed for LR white resin embedding and sectioning, followed by gold staining as described⁵³.

Human brain specimens. Fixed human brain tissue from the frontal cortex of individuals with neuropathologically verified CTE was provided from the VA-BU-SLI Brain Bank of the Boston University Alzheimer's Disease Center CTE Program, including 16 patients with a history of exposure to TBI and 8 age-matched healthy controls (Supplementary Table 1)⁶. Next of kin provided written consent for participation and brain donation. Institutional review board approval for brain donation was obtained through the Boston University Alzheimer's Disease Center, CTE Program, and the Bedford VA Hospital. Institutional review board approval for neuropathological evaluation was obtained through Boston University School of Medicine⁶. Our studies on human samples have been approved by our Institutional Review Boards at Boston University and Beth Israel Deaconess Medical Center.

Transgenic overexpression and knockout mice. Tau-transgenic mice⁵⁴ and tau-knockout mice⁵⁵ (Jackson laboratory) in the C57BL/6 background were generated, as described^{29,32,42}. Animal care and use for the experiments have been approved by Institutional Animal Care and Use Committees at Beth Israel Deaconess Medical Center.

Cell culture. Neuronal cell lines including SH-SY5Y, PC12, H4 cells (purchased originally from American Type Culture Collection) were cultured in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal calf serum. The cell lines have not been authenticated or tested for mycoplasma contamination. The media were supplemented with 100 Units ml^{-1} penicillin/streptomycin. PC12 cells were differentiated with NGF (50 ng ml^{-1}) and cultured for 2 days before stress. SY5Y

cells (2.5×10^5 per ml) were transiently co-transfected with 2.5 μ g GFP- τ , 2.5 μ g Cdk5 and 2.5 μ g p25 with Lipofectamine 2000 (Invitrogen) as described⁴⁰. Cells were treated with *cis* or *trans* mAbs at 8.0 μ g ml⁻¹ once 4 h after transfection until observation.

Cell viabilities were examined using Live & Dead cell assay kit (Abcam) according to the manufacturer. For apoptosis assay, cells were trypsinized and suspended in a binding buffer (10 mM HEPES, pH 7.4; 140 mM NaCl; 2.5 mM CaCl₂), stained with Annexin V (Biolegend 640912) for 15 min and subjected for flow cytometry. Brain or cell extracts were applied to culture SY5Y cells for 18 h and the cell viabilities were examined as described earlier. We performed the cell or brain lysate extraction using RIPA buffer but efficiently dialysed the extracts against PBS for 48 h dialysis with 4 changes of buffer.

Stably overexpressing RFP- or GFP- τ SY5Y cells were routinely generated. Briefly, the plasmids pcDNA3.1- τ -RFP and -GFP constructed through restriction sites and were transfected into SY5Y cells via Lipofectamine 2000. Cells stably expressing τ selected with G418. Equal number of GFP- and RFP- τ expressing cells (2.5×10^5 per ml) were cocultured, treated or untreated with either *cis* or *trans* mAbs at concentration of 170 μ g ml⁻¹ for 18 h before moving into hypoxia chamber.

To test cell or brain lysates, accordingly, we performed the extraction using RIPA buffer but efficiently dialysed the extracts against PBS with a cocktail of protease inhibitors at 4 °C for 48 h dialysis with 4 changes of the buffer. After dialysis, we examined *cis* τ concentration and conformation with immunoblotting and applied the amount of dialysed lysates similar to the original lysates to culture dishes. SY5Y cells were treated with 3-methyladenine at concentration of 5 mM for 24 h.

Primary neurons were prepared from 17-day-old embryonic mouse brain cerebral cortex of either sex. Neurons were seeded on pre-coated culture dishes (2×10^5 per ml). The medium was then changed to neurobasal medium supplemented with B27 (Invitrogen) and 1 mM L-glutamine as described⁵². Neurons were infected with lentivirus coding for either GFP- or RFP- τ for 72 h.

Mitochondrial transport assay. PC12 cells were differentiated with NGF (Cell Signaling) at 50 ng ml⁻¹ and cultured for 2 days before stress. They were treated with *cis* or *trans* mAbs at 85 μ g ml⁻¹ for 18 h and transferred into hypoxia chamber for 48 h more. Then, mitochondria were stained using Mitotracker Green FM (Life technologies) according to manufacturer and observed with a Zeiss confocal microscope for 30 min using an incubation chamber with 5% CO₂ at 37 °C. Fluorescent images of labelled mitochondria in the longest process of each PC12 cell were acquired at intervals of 5 s over a period of 30 min. Individual mitochondrial movements were analysed with ZEN 2008 software (Zeiss). Differences in the position of each mitochondrion between two frames during each 5 s interval were exported to Excel, and they were classified and scored as stationary, fast ($>0.05 \mu$ m s⁻¹) or slow ($<0.05 \mu$ m s⁻¹) movements.

Traumatic brain injury. The mouse TBI model was used as previously described^{25,56}. Briefly, male C57BL/6 mice (2–3 months old) obtained from the Jackson Laboratories (Bar Harbour, ME) were randomized to undergo injury or sham-injury. The mice were anaesthetized for 45 s using 4% isoflurane in a 70:30 mixture of air:oxygen. Anaesthetized mice were placed on a delicate task wiper (Kimwipe, Kimberly-Clark, Irving, TX) and positioned such that the head was placed directly under a hollow guide tube. The mouse's tail was grasped. A 54-gram metal bolt was used to deliver an impact to the dorsal aspect of the skull, resulting in a rotational acceleration of the head through the Kimwipe. Mice underwent single severe injury (ssTBI, 60-inch height), single mild injury (mTBI, 28-inch height), or repetitive mild injuries (rmTBI, 7 injuries in 9 days). Sham-injured mice underwent anaesthesia but not concussive injury. All mice were recovered in room air. Anaesthesia exposure for each mouse was strictly controlled to 45 s. Subsequent behavioural and histopathological testing was conducted in a blinded manner. Blast-induced TBI mouse model was performed as described⁵. Briefly, anaesthetized adult wild-type C57BL/6 male mice were exposed to a single blast or sham blast, removed from the apparatus, monitored until recovery of gross locomotor function, and then transferred to their home cage. Maximum burst pressure compatible with 100% survival and no gross motor abnormalities was ascertained empirically. All these and following animal experiments were approved by the Boston Children's Hospital, Beth Israel Deaconess Medical Center and/or Boston University and IACUC and complied with the NIH Guide for the Care and Use of Laboratory Animals.

Antibody treatment of mice. Mice undergoing TBI were randomized to treatment with anti-*cis* P- τ monoclonal mouse antibody or mouse IgG2b. Mice received 1 dose of *cis* antibody or IgG2b intraperitoneal pre-treatment (i.p. 200 μ g per mouse) 3 days before the injury (which was omitted in some experiments), post-injury treatment with single intracerebroventricular (ICV) treatment (20 μ g in 5 microlitres) 15 min after injury, then post-treatment 200 μ g i.p. every 4 days for 3 times and analysed brains 14 days later for

immunoblotting or fEPSP recording, followed by 200 μ g i.p. weekly for another 1.5 months (with a total 2 months of treatment) before the elevated plus maze or the Morris Water Maze in a double-blinded manner. After the above treatment, some mice received further antibody treatment, by 200 μ g i.p. biweekly for another 4 months before assaying *cis* P- τ spread, τ aggregation and tauopathy and brain atrophy at 6 months after ssTBI, as described^{29,32,42}. For all behavioural tests, experimenters were blinded to injury and treatment status, using colour-coding stored in a password protected computer.

Electrophysiology. Mice were anaesthetized with isoflurane (NDC 10019-360-40, Baxter Healthcare Corporation Deerfield, IL, USA) and decapitated. The brains were quickly removed and placed for sectioning in ice-cold treatment artificial cerebrospinal fluid (tACSF) containing (in mM) NaCl 124, KCl 3, NaH₂PO₄ 1.25, NaHCO₃ 26, CaCl₂ 2, MgSO₄ 2, and glucose 10 (pH 7.4, and bubbled with 95% O₂ and 5% CO₂ gas mixture). Cortical slices (thickness 350 μ m) were cut with a Vibratome 1000P (Leica VT1000P, Leica Microsystems Inc., Buffalo Grove, IL, USA) and transferred to a chamber with oxygenated tACSF for 90 min at 30 °C before recording.

Field excitatory postsynaptic potentials (fEPSP) were recorded using a multi-electrode array recording system (MED64 system) with MED-P5155 probe (AutoMate Scientific, Inc., Berkeley, CA, USA) in this study. After incubation, one cortical slice was positioned in the centre of the MED64 probe (to fully cover the 8 × 8 electrode array) with oxygenated recording ACSF (rACSF) containing (in mM) NaCl 124, KCl 3, NaH₂PO₄ 1.25, NaHCO₃ 26, CaCl₂ 2, MgSO₄ 1, and glucose 10 (pH 7.4) at 30 °C. A fine nylon mesh and a mesh anchor were placed on top of the slice to immobilize the slice during recording. The probe with immobilized slice was connected to two MED64 amplifiers (MED64 Head Amplifier (MED-A64HE1) and Main Amplifier (MED-A64MD1), AutoMate Scientific, Inc., Berkeley, CA, USA). The slice was continuously perfused with oxygenated, fresh rACSF at the rate of 2 ml min⁻¹ using a peristaltic pump (Minipuls 3, Gilson Inc., Middleton, WI).

Data was collected using Mobius software (Mobius 0.4.2). Field potentials were induced by single pulses (0.2 ms) delivered at 0.05 Hz through one planar micro-electrode in layer V of cortical slice. We used stimulus intensity sufficient to induce a 50% of the maximal fEPSP slope in all experiments. The fEPSP was recorded from the channels in layer II/III. A stable fEPSP slope for at least 20 min was recorded as baseline. The induction protocol of LTP that we used is 5 Hz theta burst (each burst consists of 4 pulses at 100 Hz). The data were filtered at 10 kHz and digitized at a 20 kHz sampling rate. Data were analysed off line by the MED64 Mobius software. For quantifying the level of LTP, the mean of fEPSP slope (10–40%) within the last 10 min of recording was normalized and expressed as a fold change of the averaged baseline (first 10 min of the baseline). Three successive responses were averaged. Statistics were performed using the number of slices as 'n' value, and one to two slices per animal. P values were calculated using one-way ANOVA with Bonferroni post hoc test.

Morris water maze. A Morris water maze (MWM) paradigm was used to evaluate spatial learning and memory as described^{25,57}. Briefly, a white pool (83 cm diameter, 60 cm deep) was filled with water to 29 cm depth. Water temperature was maintained at approximately 24 °C. Several highly visible intra- and extra-maze cues were located in and around the pool. The target platform (a round, clear, plastic platform 10 cm in diameter) was positioned 1 cm below the surface of the water. During hidden and visible platform trials, mice were randomized to one of four starting quadrants. Mice were placed in the tank facing the wall and given 90 s to find the platform, mount the platform, and remain on it for 5 s. Mice were then placed under a heat lamp to dry before their next run. Time until the mouse mounted the platform (escape latency) was measured and recorded. Mice that failed to mount the platform within the allotted time (90 s) were guided to the platform by the experimenter and allowed 10 s to become acquainted with its location. Each mouse was subjected to a maximum of two trials per day, each consisting of four runs, with a 45-min break between trials. For visible platform trials, a red reflector was used to mark the top of the target platform. For probe trials, mice were placed in the tank with the platform removed and given 60 s to explore the tank. Noldus Ethovision 9 software tracked swim speed, total distance moved, and time spent in the target quadrant where the platform was previously located. When mice underwent repeat MWM testing, 2 to 3 months or 6 months after their final injury, the platform was moved to a different quadrant than that used previously.

Elevated plus maze. The elevated plus maze was used to assess anxiety/risk-taking behaviour two months after injury and carried out as described⁵⁸. Briefly, the elevated plus maze consists of two open and two closed arms (30 × 5 cm) extended out opposite from each other from a central platform (decision zone) to create a plus shape. The entire apparatus is raised 85 cm above the floor (Lafayette Instruments). Mice are placed on the centre platform of the maze, facing a closed arm, and allowed to explore the apparatus for

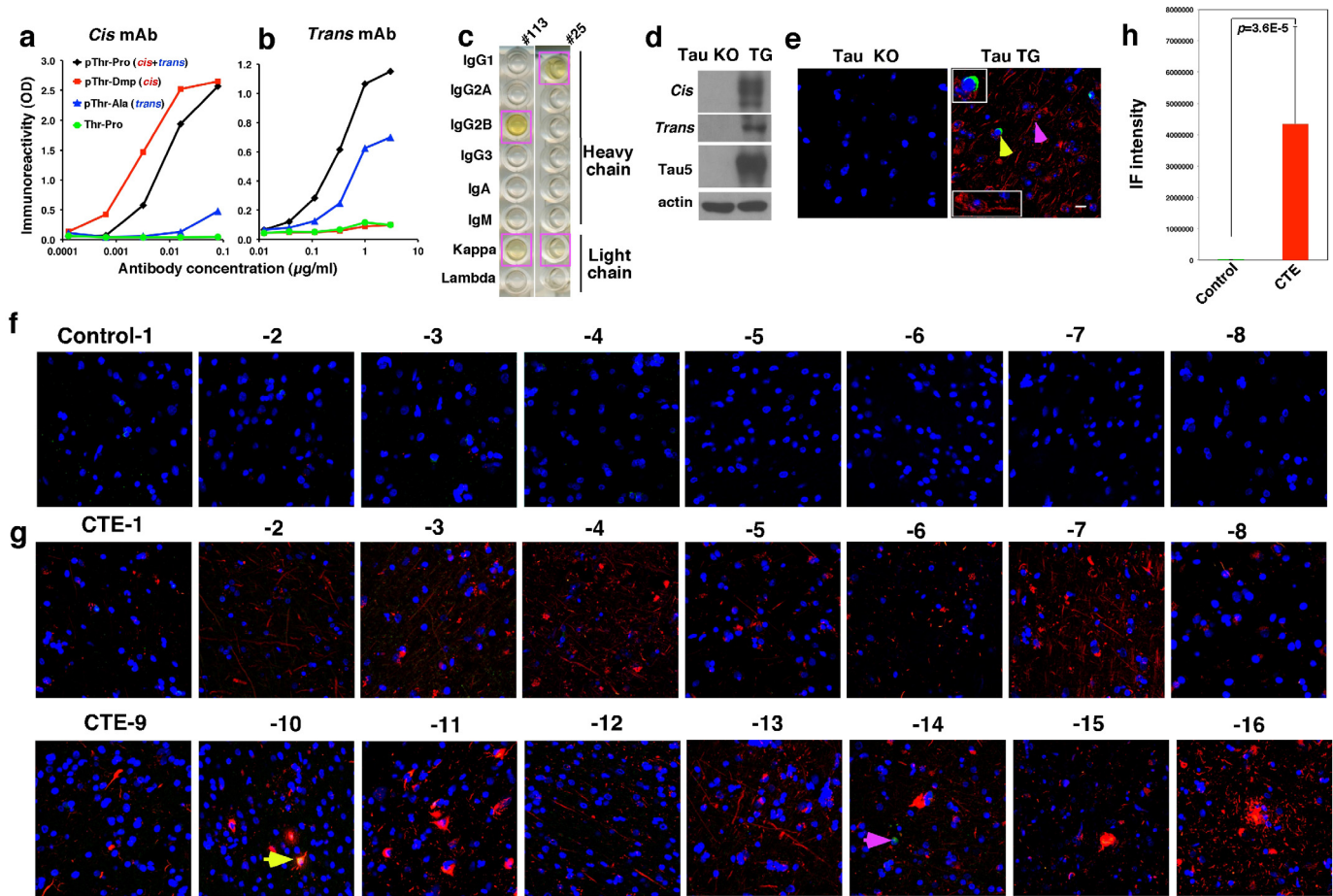
5 min. The maze is cleaned between subjects with a weak ethanol solution and dried. A computer-assisted video-tracking system (Noldus Ethovision) recorded the total time spent in the open centre (decision zone), and the two closed or 'safe' arms and the two open or 'aversive' arms. The percent time spent in the open arms is used as a surrogate measure of anxiety/risk-taking behaviour; mice with lower levels of anxiety/risk-taking behaviour spend less time in the open arms.

Immunohistochemistry. Mice were intracardially perfused with 4% para-formaldehyde at various time points after injury and brains were collected for histopathological outcomes. Serial 20 µm coronal frozen sections from sham and injured brains were cut on a cryostat (Leica) from the anterior frontal lobes through the posterior extent of the dorsal hippocampus. Every 10th section was collected and mounted on slides.

Statistical analysis. Experiments were routinely repeated at least three times, and the repeat number was increased according to effect size or sample variation. We estimated the sample size considering the variation and mean of the samples. No animals or samples were excluded from any analysis. Animals were randomly assigned groups for *in vivo* studies and for mAb treatment experiments in mice, group allocation and outcome assessment were also done in a double blinded manner. For all behavioural tests, experimenters were blinded to injury and treatment status, using colour coding stored in a password protected computer. All data are presented as the means \pm s.d. except behavioural tests where data are

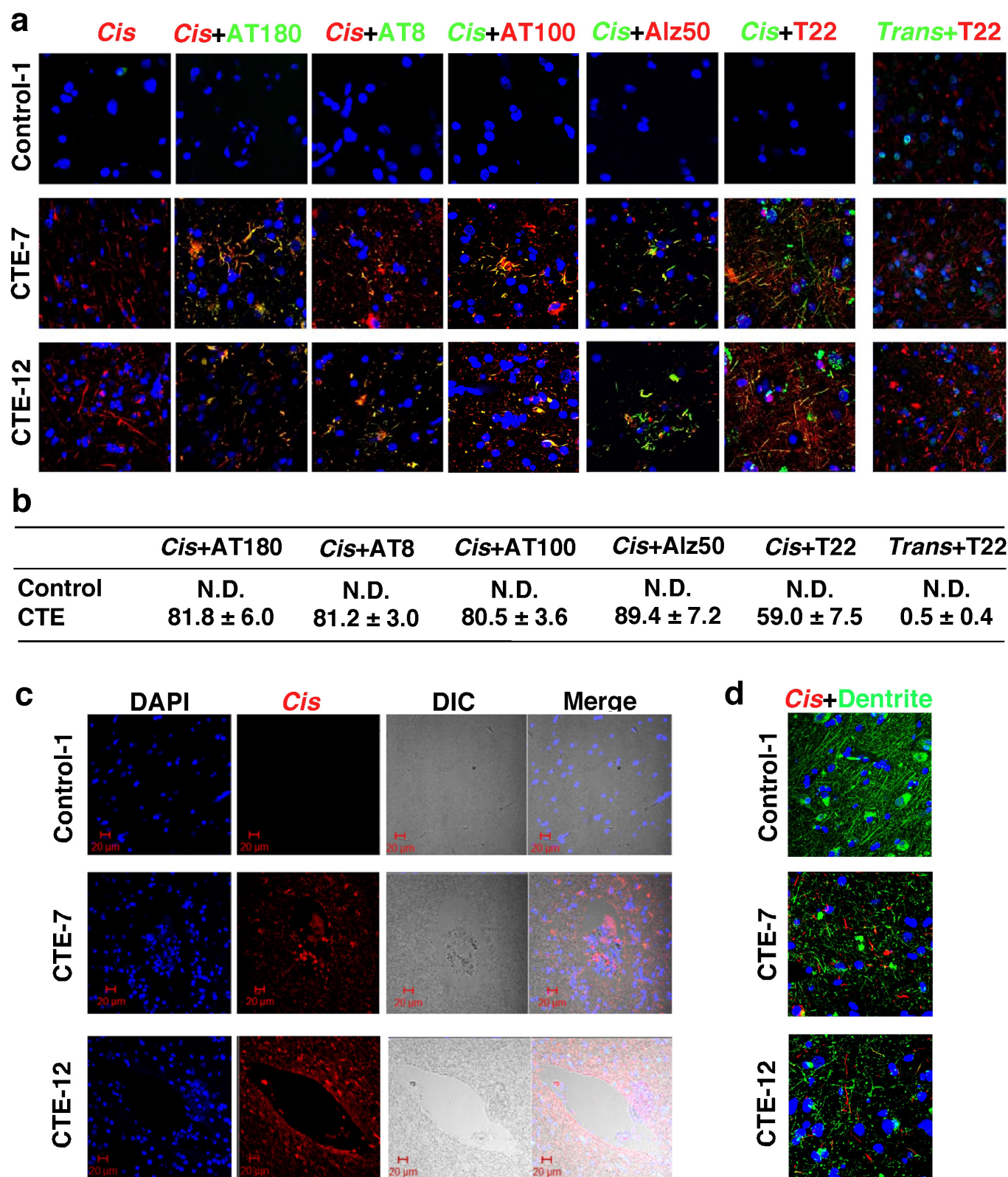
presented as the means \pm s.e.m., followed by determining significant differences using the two-tailed Student's *t* test for quantitative variables or ANOVA test for continuous or three or more independent variables or one-way ANOVA with Bonferroni post hoc test, and significant *P* values <0.05 are shown.

51. Tokuoka, H. *et al.* Brain-derived neurotrophic factor-induced phosphorylation of neurofilament-H subunit in primary cultures of embryo rat cortical neurons. *J. Cell Sci.* **113**, 1059–1068 (2000).
52. Shahpasand, K. *et al.* Regulation of mitochondrial transport and inter-microtubule spacing by tau phosphorylation at the sites hyperphosphorylated in Alzheimer's disease. *J. Neurosci.* **32**, 2430–2441 (2012).
53. Farah, C. A. *et al.* Tau interacts with Golgi membranes and mediates their association with microtubules. *Cell Motil. Cytoskeleton* **63**, 710–724 (2006).
54. Ishihara, T. *et al.* Age-dependent emergence and progression of a tauopathy in transgenic mice overexpressing the shortest human tau isoform. *Neuron* **24**, 751–762 (1999).
55. Dawson, H. N. *et al.* Inhibition of neuronal maturation in primary hippocampal neurons from tau deficient mice. *J. Cell Sci.* **114**, 1179–1187 (2001).
56. Meehan, W. P. III, Zhang, J., Mannix, R. & Whalen, M. J. Increasing recovery time between injuries improves cognitive outcome after repetitive mild concussive brain injuries in mice. *Neurosurgery* **71**, 885–892 (2012).
57. Walf, A. A. & Frye, C. A. The use of the elevated plus maze as an assay of anxiety-related behavior in rodents. *Nature Protocols* **2**, 322–328 (2007).
58. Morris, R. Developments of a water-maze procedure for studying spatial learning in the rat. *J. Neurosci. Methods* **11**, 47–60 (1984).



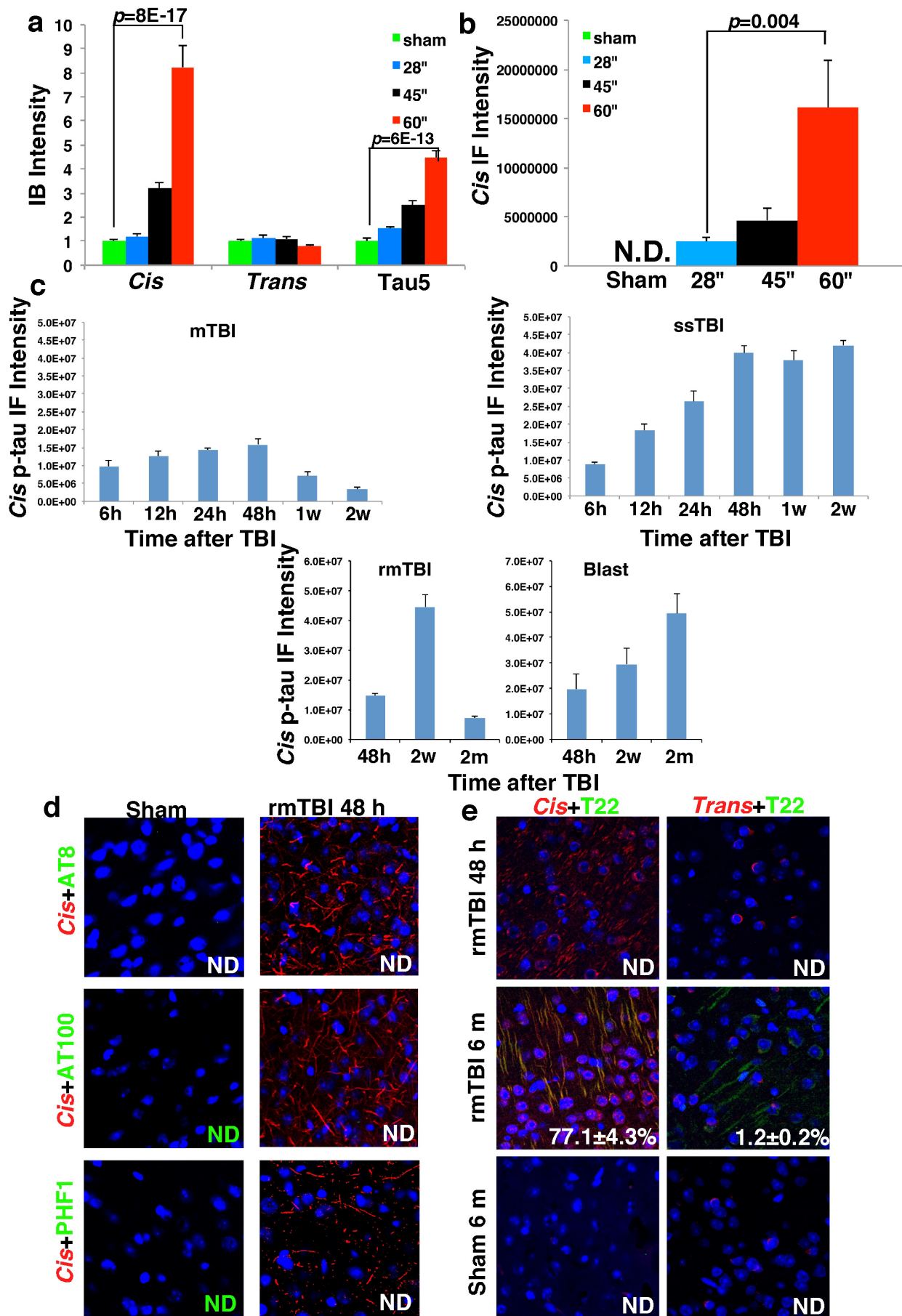
Extended Data Figure 1 | Characterization of *cis* and *trans* P-tau mAbs and robust *cis* P-tau in human CTE brains. **a, b,** Characterization of the specificity of *cis* and *trans* P-tau mAbs by ELISA. *Cis* (**a**) and *trans* (**b**) antibodies at various concentrations were incubated with *cis* (pT231-Dmp), *trans* (pT231-Ala), *cis* + *trans* (pT231-Pro) or T231-Pro tau peptides, followed by detecting the binding by ELISA. Representative examples of ELISA are shown from 3 independent experiments. pT231-Pro, CKKVAVVRpT(Pro)PKSPSSAK; pT231-Pip, CKKVAVVRpT(homoproline)PKSPSSAK; pT231-Ala, KVAVVRpT(alanine)PKSPS; pT231-Dmp (KVAVVRpT(5,5-dimethyl-L-proline)PKSPS). **c,** Determination of the isotypes of *cis* and *trans* P-tau mAbs. Isotypes of *cis* and *trans* mAb heavy and light chains were determined by ELISA assay using a commercially available assay kit. **d, e,** Characterization of the specificity of *cis* and *trans* P-tau mAbs by immunoblotting and immunofluorescence. Brain lysates (**d**) or sections (**e**) prepared from tau-deficient (KO)

or wild-type tau-overexpressing (TG) mice were subjected to immunoblotting or immunofluorescence with *cis* and/or *trans* antibody. The *cis* and *trans* signals were readily detected in TG, but not at all in KO mouse brains, with *cis* in the soma and neurites (pink arrow), but *trans* only in the soma (yellow arrow) (insets). Similar results were observed in at least three different animals. *Cis*, red; *trans*, green; DNA, blue. **f–h,** Robust *cis* P-tau in human CTE brains. 16 CTE brain tissues and 8 healthy controls were subjected to immunofluorescence, with one representative image from each case being shown (**f, g**). Yellow arrow points to a neuron expressing both *cis* (red) and *trans* (green) P-tau, while pink one to a neuron expressing only *trans* in the soma. Fluorescence immunostaining intensity of *cis* P-tau was quantified using Volocity 6.3 from Perkin Elmer (**h**). The results are expressed as means \pm s.d. and *P* values determined using the Student's *t*-test.



Extended Data Figure 2 | Colocalization of *cis* P-tau with other tau epitopes and its concentration near blood vessels in CTE brains. **a, b,** Colocalization of *cis* P-tau with other tau epitopes in CTE brains. CTE brain tissues and healthy controls were stained with *cis* mAb and AT180, AT8, AT100, Alz50 or T22 antibodies, or *trans* mAb and T22 antibodies, with two examples being shown (**a**), and then quantified their colocalization using Coloc 2, with the results being expressed in a percentage (mean ± s.d.) (**b**). N.D., not detectable.

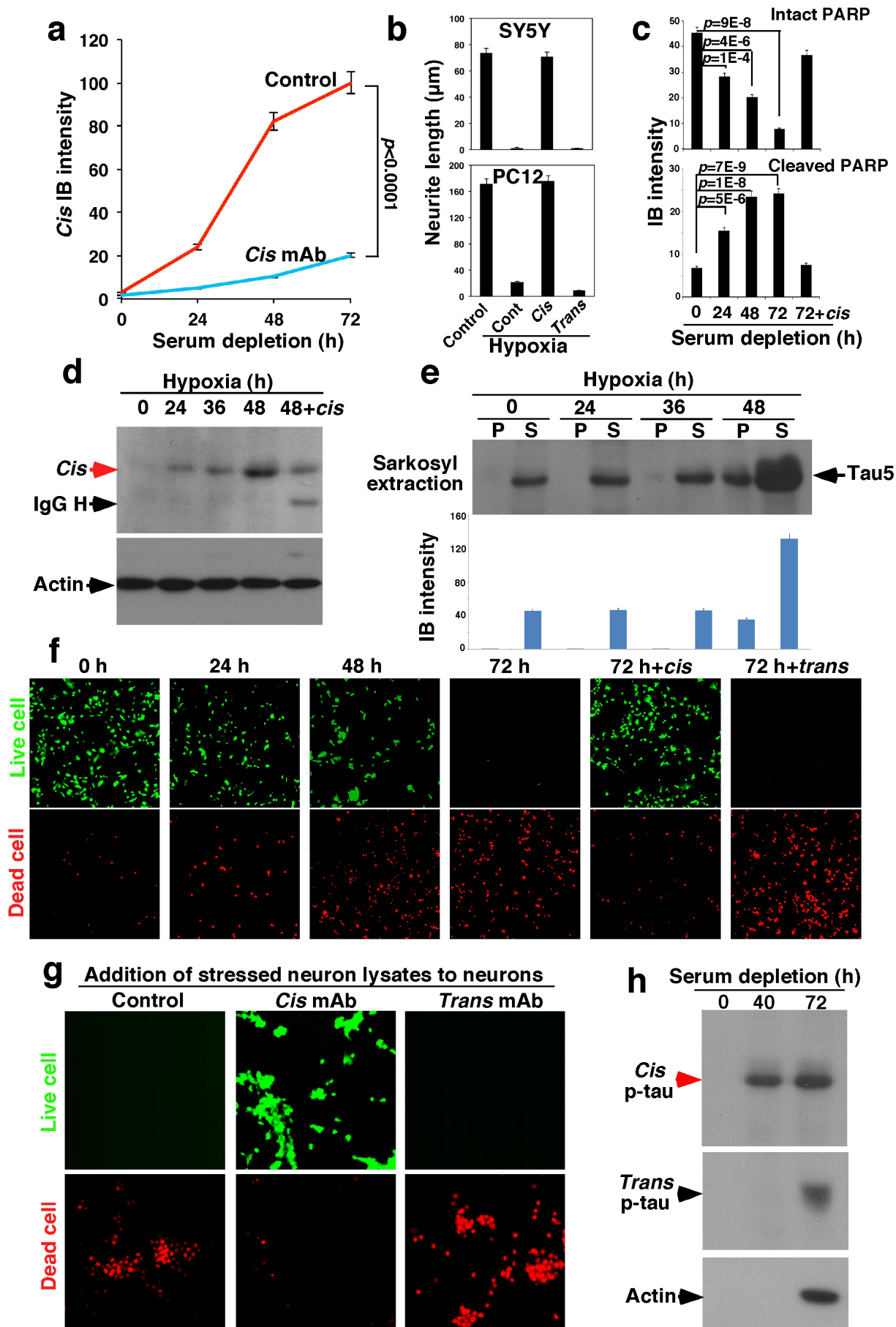
c, CTE brain tissues and healthy controls were stained with *cis* mAb, with two examples being shown. *Cis* is more prominent near blood vessels, which corresponds to the typical perivascular distribution of P-tau in CTE. **d,** CTE brain tissues and healthy controls were stained with *cis* mAb (red) and the dendritic marker MAP2 (green), along with DNA dye (blue). Colours in the text correspond to their fluorescence labels. *n* = 4.



Extended Data Figure 3 | TBI induces *cis* P-tau in a severity- and time-dependent manner long before other known tauopathy epitopes.

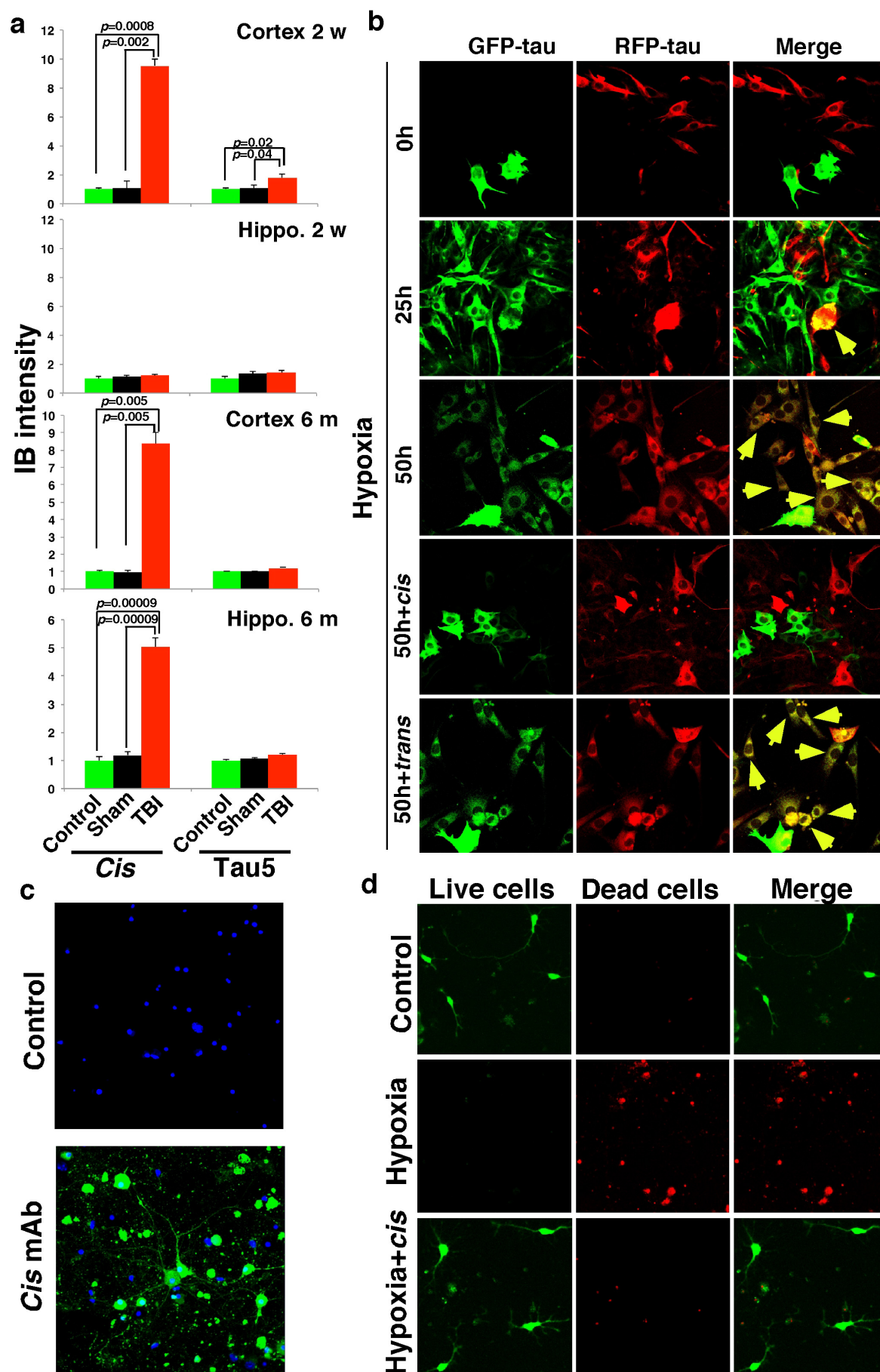
a–c, Severity- and time-dependent induction of *cis* P-tau after TBI. Quantification results of Fig. 2a–f. **d**, Robust *cis* P-tau signals are detected in neurons 48 h after rmTBI without any other tangle-related tau epitopes. 48 h after rmTBI, brain sections were stained with *cis* mAb (red) and AT8, AT100 or PHF1 (green). **e**, Robust *cis* P-tau signals are detected in neurons 48 h after rmTBI without tau oligomerization, which appear and colocalize with *cis* P-tau

at 6 months after TBI. 48 h or 6 months after rmTBI or sham treatment, brain sections were immunostained with T22 (green) and *cis* or *trans* mAb (red). The results in 48 h sham mice were similar to those at 6 months (data not shown). The colocalization of red and green signals was quantified using Coloc-2, with the results being shown in percentages. ND, not detectable. $n = 3–4$. The results are expressed as means \pm s.d. and P values determined using the Student's t -test.



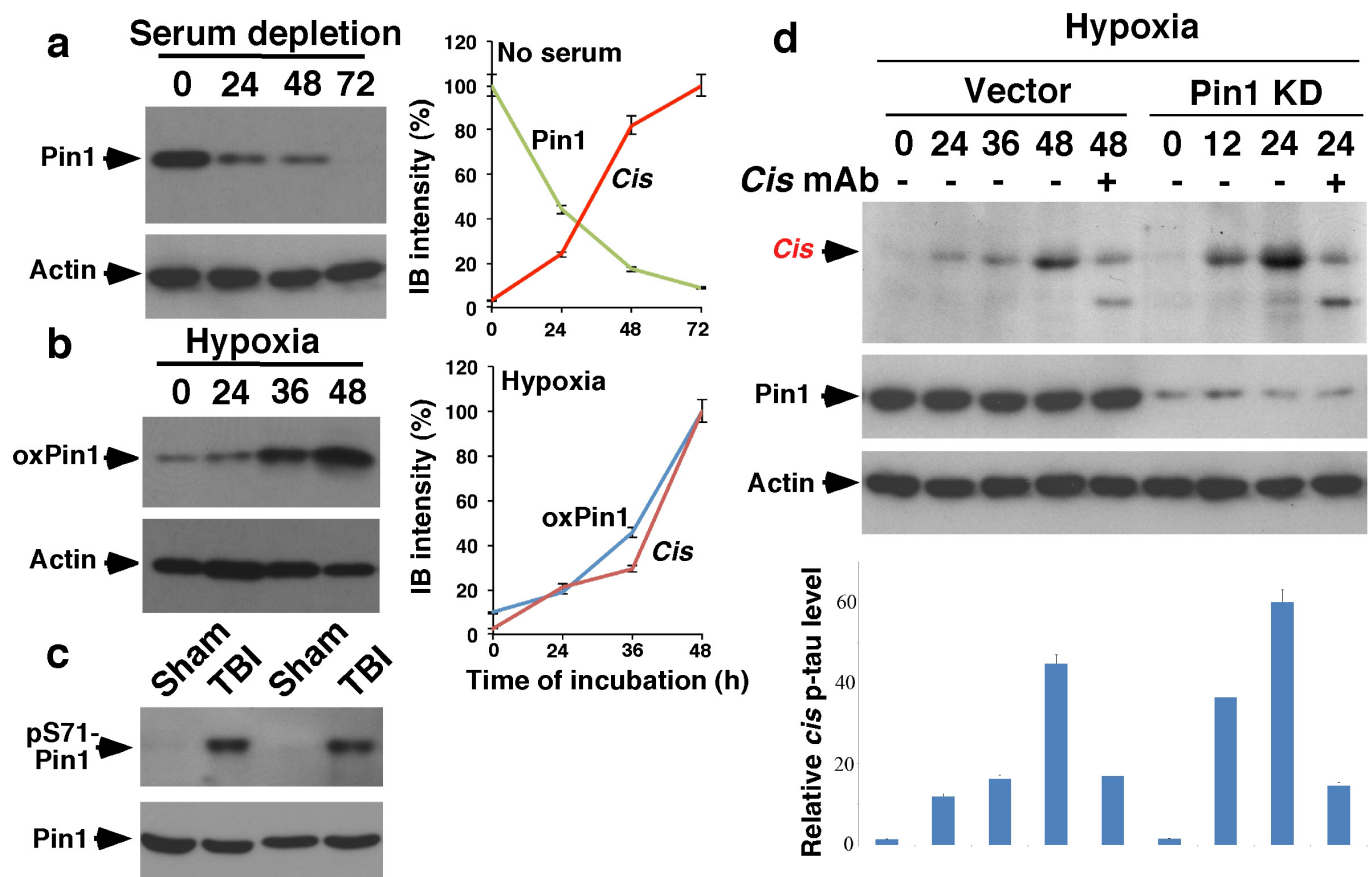
Extended Data Figure 4 | Stressed neurons robustly produce *cis* P-tau, *cis* P-tau is released from stressed neurons and neurotoxic, but is effectively blocked by *cis*, but not *trans*, mAb. **a–c,** Quantification results of Fig. 4a, b and f, respectively. The results are expressed as means \pm s.d. and *P* values determined using the two-way ANOVA test (**a**) and Student's *t*-test (**c**). **d,** Hypoxia induces *cis* P-tau, which is blocked by *cis* mAb. SY5Y neurons expressing a control vector were cultured in the hypoxia chamber in the absence or presence of *cis* or *trans* mAb for the times indicated, followed by immunoblotting for *cis* P-tau. **e,** Hypoxia induces *cis* P-tau before tau aggregation. SY5Y neurons were subjected to hypoxia for the times indicated, followed by sarkosyl extraction before immunoblotting with Tau5 mAb and quantification. **f,** Hypoxia induces cell death, which are blocked by *cis*, but not

trans, mAb. SY5Y neurons were cultured in the hypoxia chamber in the absence or presence of *cis* or *trans* mAb for the times indicated, followed by live and dead cell assay using the LIVE/DEAD Viability/Cytotoxicity Kit. **g,** Stressed neuron lysates are neurotoxic, which are neutralized by *cis*, but not *trans*, mAb. Cell lysates were prepared from stressed SY5Y neurons and then added to growing SY5Y neurons directly (Control) or after immunodepletion with *cis* or *trans* mAb to remove *cis* or *trans* P-tau, respectively for 3 days, followed by live and dead cell assay. **h,** *Cis* P-tau is released from stressed neurons. SY5Y neurons were cultured in the absence of serum for the times indicated and culture media were collected and centrifuged, followed by analysing the supernatants for *cis* and *trans* P-tau with actin as an indicator of cell lysis.



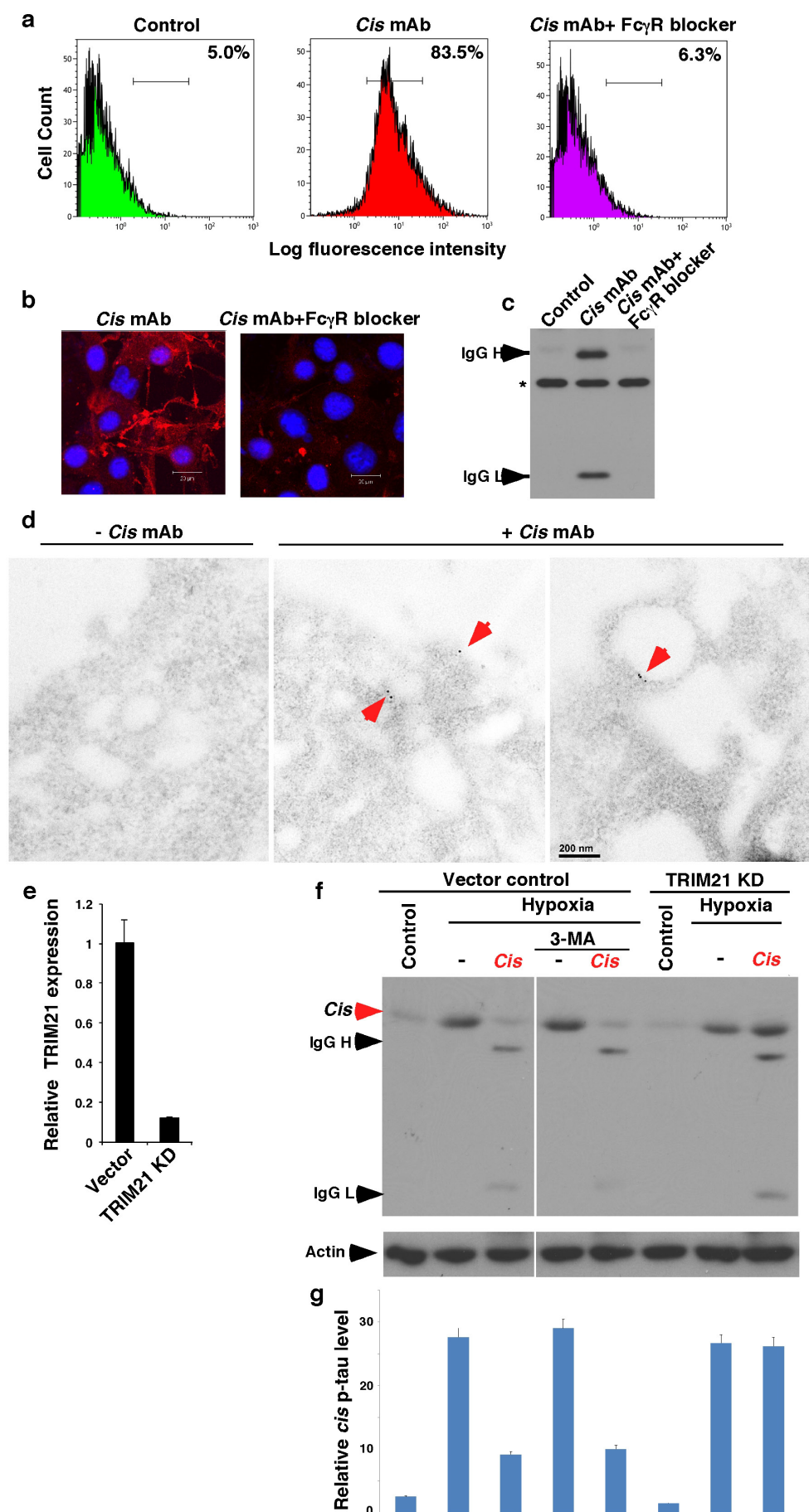
Extended Data Figure 5 | *Cis* P-tau spreads after rmTBI or neuronal stress, and hypoxia induces cell death in primary neurons, which is blocked by *cis* mAb. **a**, *Cis* P-tau spreads in the brain after rmTBI. Quantification results of Fig. 3c. **b**, *Cis* P-tau spreads after neuronal stress. GFP-tau or RFP-tau SY5Y neurons were co-cultured and subjected to hypoxia or control treatment in the presence or absence of *cis* or *trans* mAb for different times, followed by assaying cells expressing both GFP-tau and RFP-tau (arrows) to determine tau spreading among cells. The results are expressed as means \pm s.d. and *P* values

determined using the Student's *t*-test. **c**, *Cis* mAb enters primary neurons. Primary neurons were established from mouse embryos and differentiated *in vitro* and *cis* mAb was added to culture media, followed by immunostaining with secondary antibodies. **d**, Hypoxia induces cell death in primary neurons, which is effectively blocked by *cis* mAb. Primary neurons were cultured in the hypoxia chamber in the absence or presence of *cis* mAb for 48 h, followed by live (green) and dead (red) cell assay using the LIVE/DEAD Viability/Cytotoxicity Kit.



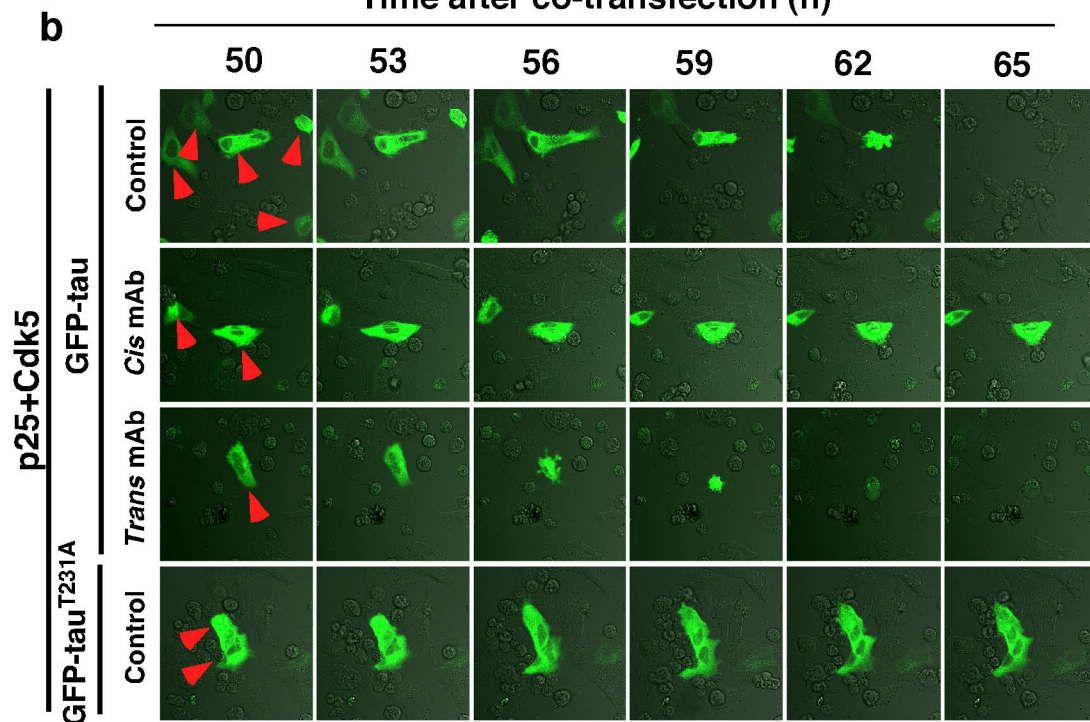
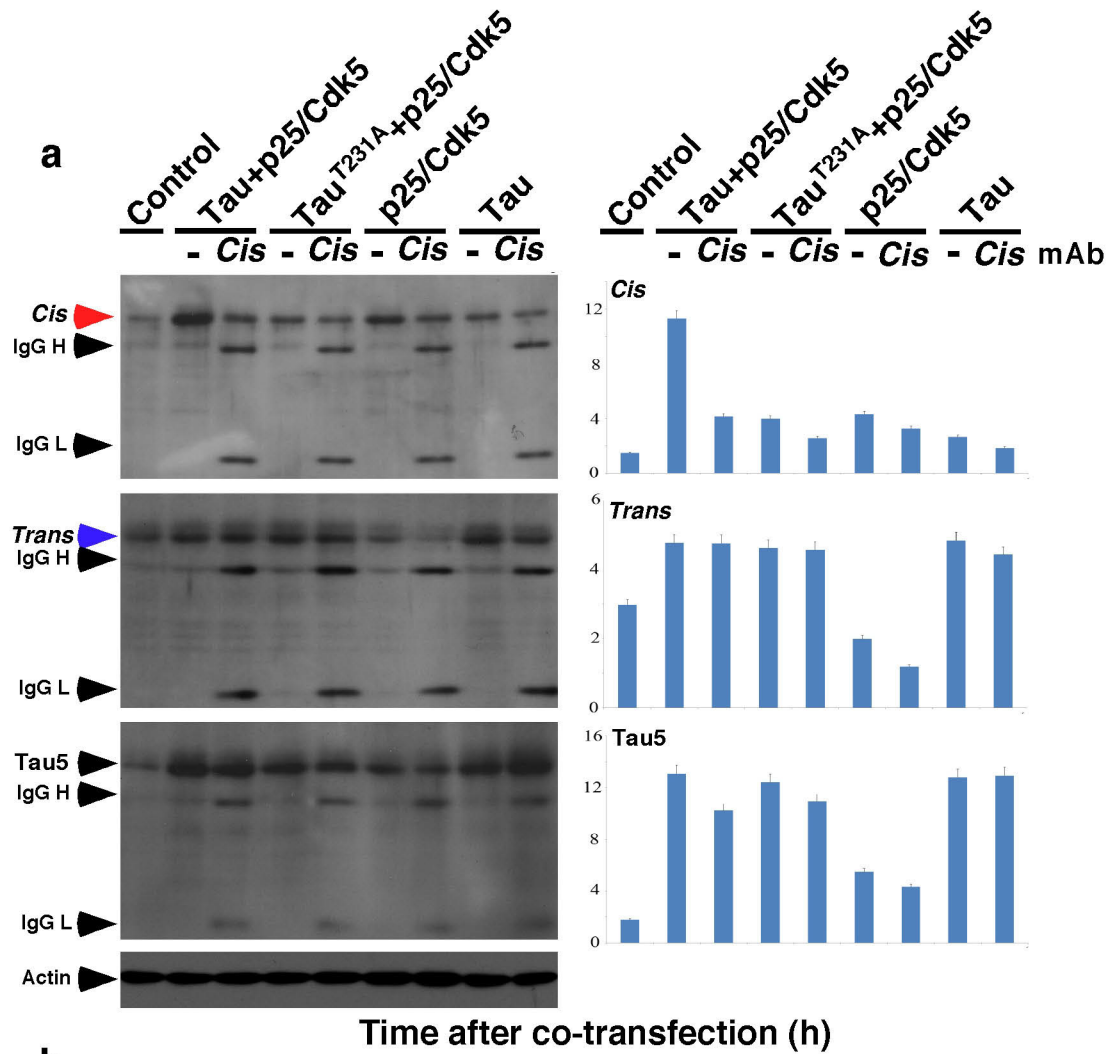
Extended Data Figure 6 | Pin1 inhibition by multiple mechanisms contributes to *cis* P-tau induction after neuron stress and TBI. **a**, Pin1 is downregulated and correlates with *cis* P-tau induction after serum starvation. Cells were subjected to serum starvation for times indicated, followed by immunoblotting, with the right panel showing the correlation of Pin1 down regulation with *cis* P-tau induction from Fig. 4a. **b**, Pin1 is oxidized and correlates with *cis* P-tau induction after hypoxia. SY5Y cells were subjected to hypoxia for times indicated, followed by immunoblotting for C113 oxidized

Pin1, with the right panel showing the correlation of Pin1 oxidization with *cis* P-tau induction from Extended Data Fig. 6d. **c**, Pin1 is inhibited in TBI mouse brains. Mouse brains 48 h after ssTBI were subjected to immunoblotting and quantification for Pin1 and S71 phosphorylated Pin1. **d**, Pin1 knockdown potentiates the ability of hypoxia to induce *cis* P-tau. Pin1-knockdown or vector control SY5Y cells were subjected to hypoxia treatment for the times indicated in the presence or absence of *cis* mAb, followed by immunoblotting and quantification for *cis* P-tau levels. The results are expressed as means \pm s.d.



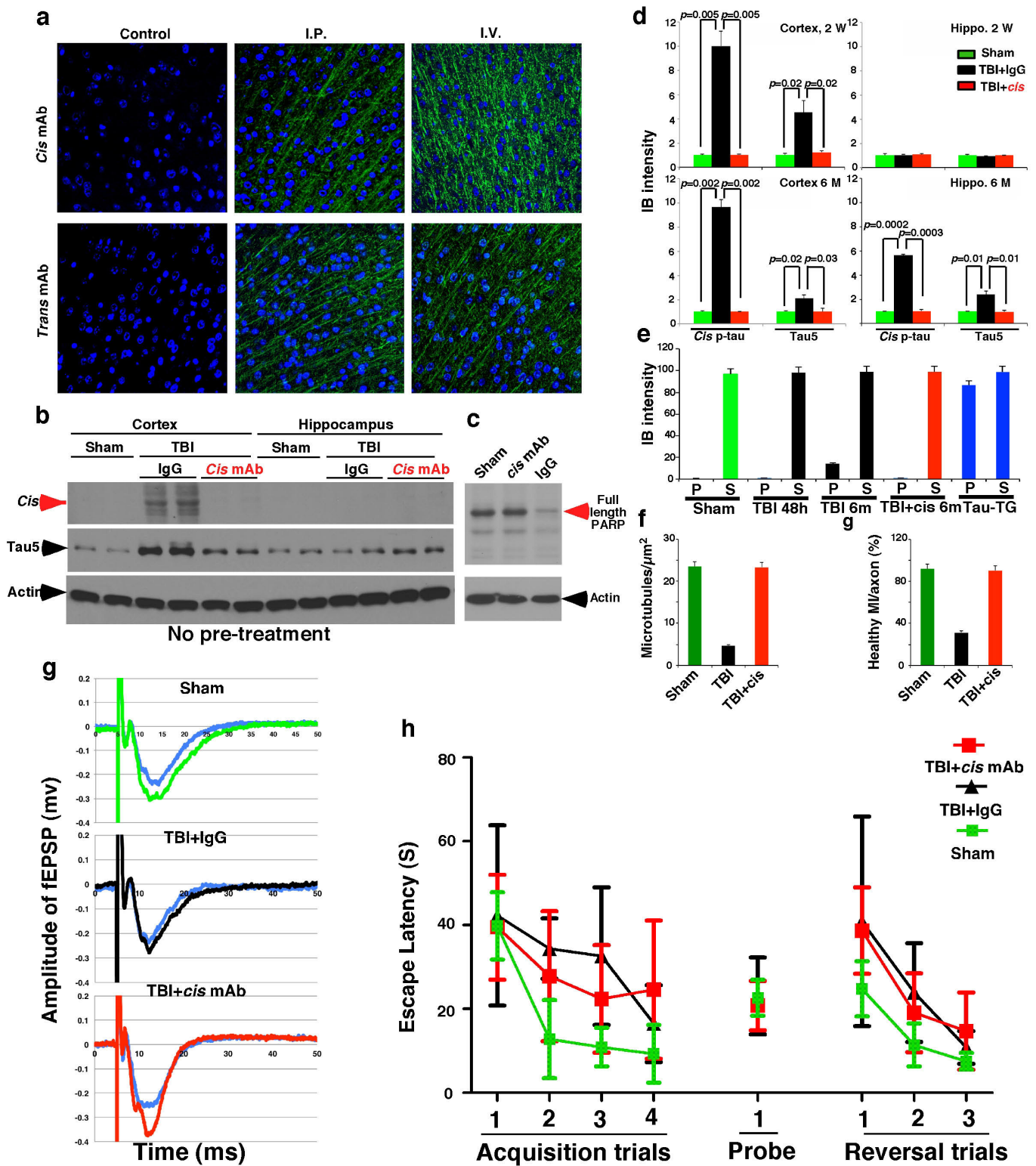
Extended Data Figure 7 | Inhibition of Fc γ R binding blocks *cis* mAb from entering neurons and TRIM21 KD fully prevented *cis* antibody from ablating *cis* P-tau in neurons. a–d, Inhibition of Fc γ R binding potentially blocks *cis* mAb from entering neurons. *Cis* mAb was added to neurons in the absence or presence of a human Fc γ R-binding inhibitor, followed by detecting the binding of *cis* mAb to the cell surface by FACS (a), entry of *cis* mAb into cells by immunofluorescence (b), immunoblotting (c) and electron microscopy after immunogold labelling (d). The FcR binding inhibitor fully blocked *cis* mAb from binding to the cell surface and entering neurons. Electron microscopy showed that *cis* mAb bound to the cell surface and endocytic

vesicles (red arrows). **e, f**, TRIM21 knockdown fully prevents *cis* antibody from ablating *cis* P-tau in neurons. TRIM21 was stably knocked down in SY5Y neuronal cells using a validated TRIM21 shRNA lentiviral vector and confirmed by real-time RT–PCR analysis of TRIM21 mRNA expression (e). TRIM21 knockdown or vector control SY5Y cells were subjected to hypoxia treatment in the presence or absence of *cis* mAb and/or 3-methyladenine, an autophagy inhibitor, followed by immunoblotting, followed by quantifying *cis* P-tau levels normalized actin levels (lower panel) (f). The results are expressed as means \pm s.d.



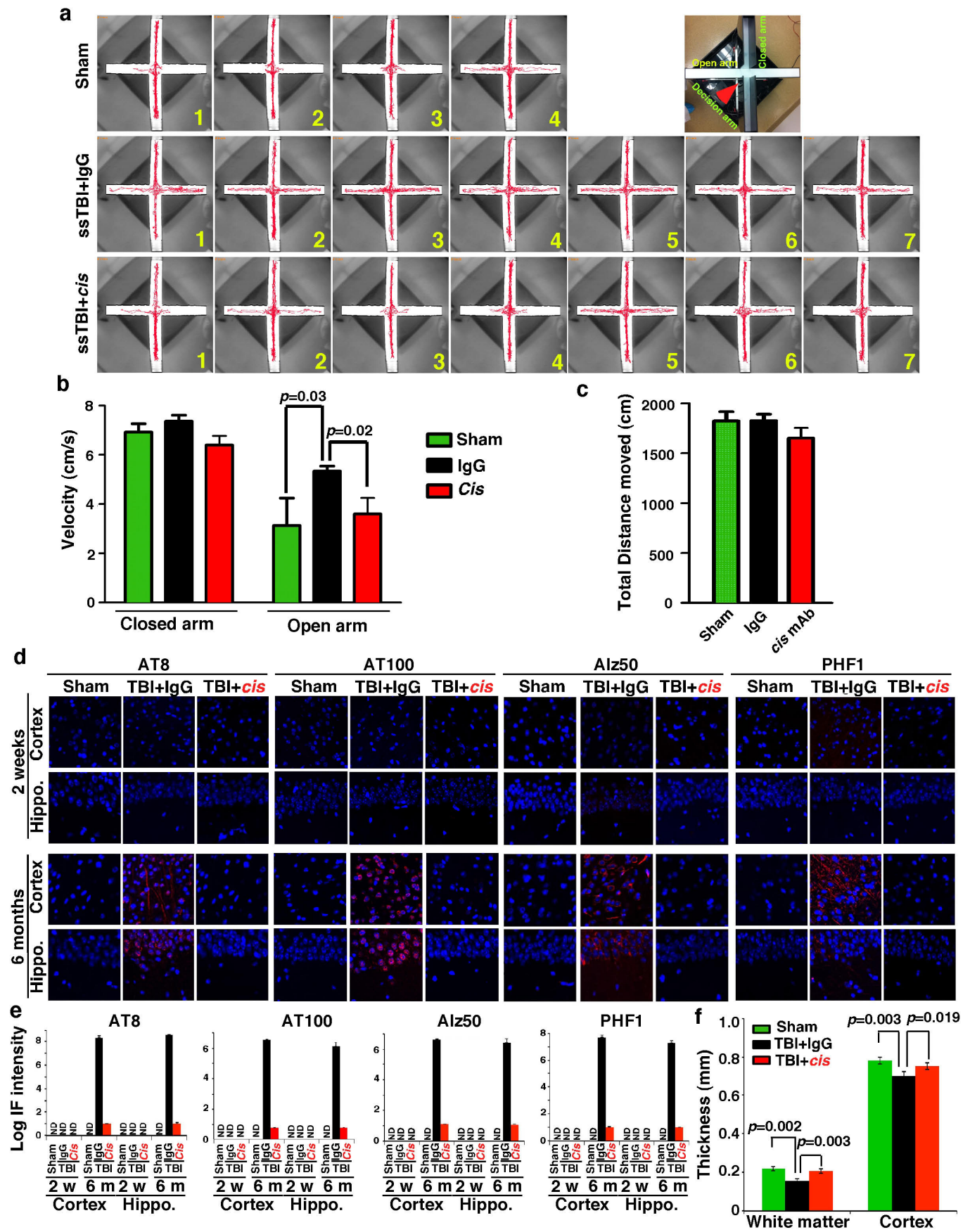
Extended Data Figure 8 | *Cis* pT231-tau is both necessary and sufficient for P-tau to induce neuronal cell death *in vitro*. **a**, SY5Y cells were co-transfected with non-tagged indicated constructs in the absence and presence of *cis* mAb followed by immunoblotting with quantification on the right panel.

b, SY5Y cells were co-transfected with GFP-tau, or GFP-tau(T231A) and p25/Ckd5 in the absence and presence of *cis* or *trans* mAb followed by live-cell confocal video (see Supplementary Videos 5, 6). Red arrows point to GFP-tau or-tau(T231A) expressing cells. The results are expressed as means \pm s.d.



Extended Data Figure 9 | *Cis* mAb effectively blocks *cis* P-tau induction and spread, tau aggregation, and restores neuronal ultrastructures, apoptosis and defective LTP after TBI. **a**, Peripherally administrated *cis* and *trans* mAbs enter neurons in brains. 250 µg of biotinylated *cis* or *trans* mAb was injected intraperitoneally or intravenously into B6 mice, followed by detecting the biotinylated *cis* mAb in brains 3 days later. **b, c**, *Cis* mAb effectively blocks *cis* pT231-tau induction and apoptosis. ssTBI mice were randomly and blindly treated with *cis* mAb or IgG isotype control, i.c.v. (intracerebroventricular) 20 µg per mouse 15 min after injury, and then i.p. 200 µg every 4 days for 3 times, followed by subjecting brains to immunoblotting for *cis* P-tau (**b**) and PARP cleavage (**c**), with sham as controls. **d–f**, *Cis* mAb effectively blocks *cis* pT231-tau induction and spread, tau aggregation and restores neuronal ultrastructures. ssTBI mice in **c–f** received additional i.p. 200 µg per mouse 3 day before injury. **d**, Quantification of immunoblotting in Fig. 5a. **e**, Quantification of immunoblotting in Fig. 5b. **f**, Quantification of electron microscopy images in Fig. 5c. $n = 3$. The results are expressed as means \pm s.d. and P values determined using Student's t -test. **g**, *Cis* mAb treatment of ssTBI mice rescues defective LTP in the

cortex. fEPSPs were recorded in the layer II/III by stimulating the vertical pathway (the layer V to II/III) in the cortex. Robust LTP was induced by 5 Hz theta-burst in the cortical slices of sham mice ($n = 15$ slices, 9 mice), but was deficient in the cortex of IgG-treated TBI mice ($n = 9$ slices, 5 mice). However, LTP magnitude was restored to the control level in *cis* mAb-treated TBI animals ($n = 9$ slices, 5 mice). The representative recordings were presented. **h**, No significant effects of *cis* pT231-tau mAb treatment on Morris Water Maze performance. 8 weeks after ssTBI, mice underwent Morris Water Maze (MWM) testing consisting of 4 acquisition trials (hidden platform) daily for 4 days (4 runs per trial), a probe trial, followed by a 3 reversal trials (hidden platform) daily for 3 days. Compared to sham mice, injured mice demonstrated increased latency to find the hidden platform in acquisition and reversal trials ($P < 0.001$). There was no difference in injured *cis* mAb mice compared to injured IgG treated mice in acquisition trials ($P = 0.5$) or reversal trial ($P = 0.9$). For probe trials, injured mice performed similarly to sham mice ($P = 0.7$) and injured *cis* mAb treated mice performed similarly to injured IgG treated mice ($P = 0.2$). $n = 4–7$. The results are expressed as means \pm s.e.m. and P values determined using ANOVA.



Extended Data Figure 10 | *Cis* mAb treatment effectively restores risk-taking behaviour and prevents tauopathy development and spread as well as brain atrophy after TBI. **a–c,** *Cis* mAb treatment effectively restores risk-taking behaviour 2 months after ssTBI. Video-tracking data of each of all mice shows that ssTBI mice treated with *cis* mAb ($n = 7$) spent similar and very little time in the open arm compared to sham mice ($n = 4$), but much less time than TBI mice treated with IgG2b ($n = 7$) (**a**). *Cis* mAb-treated ssTBI mice had similar performance to sham in travelling velocity, but IgG2b-treated ssTBI mice travelled a greater velocity in the open arm (**b**). All three groups travelled

similar total distance (**c**). Results are expressed as mean \pm S.E.M. and P values determined using the Student's t -test. **d–f,** *Cis* mAb treatment effectively prevents tauopathy development and spread as well as brain atrophy 6 months after ssTBI. ssTBI mice were treated with *cis* mAb or IgG control for 2 weeks or 6 months, with sham mice as controls, followed by immunofluorescence with various tauopathy epitopes (**d**), with immunostaining fluorescence intensity in the cortex and hippocampus being quantified (**e**), or to NeuN immunostaining for determining the thickness of the cortex and white matter at 6 months after TBI (**f**). $n = 4$.

Small-scale filament eruptions as the driver of X-ray jets in solar coronal holes

Alphonse C. Sterling¹, Ronald L. Moore^{1,2}, David A. Falconer^{1,2} & Mitzi Adams¹

Solar X-ray jets are thought to be made by a burst of reconnection of closed magnetic field at the base of a jet with ambient open field^{1,2}. In the accepted version of the ‘emerging-flux’ model, such a reconnection occurs at a plasma current sheet between the open field and the emerging closed field, and also forms a localized X-ray brightening that is usually observed at the edge of the jet’s base^{1,3}. Here we report high-resolution X-ray and extreme-ultraviolet observations of 20 randomly selected X-ray jets that form in coronal holes at the Sun’s poles. In each jet, contrary to the emerging-flux model, a miniature version of the filament eruptions that initiate coronal mass ejections^{4–7} drives the jet-producing reconnection. The X-ray bright point occurs by reconnection of the ‘legs’ of the minifilament-carrying erupting closed field, analogous to the formation of solar flares in larger-scale eruptions. Previous observations have found that some jets are driven by base-field eruptions^{8–11}, but only one such study, of only one jet, provisionally questioned the emerging-flux model¹². Our observations support the view that solar filament eruptions are formed by a fundamental explosive magnetic process that occurs on a vast range of scales, from the biggest mass ejections and flare eruptions down to X-ray jets, and perhaps even down to smaller jets that may power coronal heating^{10,13,14}. A similar scenario has previously been suggested, but was inferred from different observations and based on a different origin of the erupting minifilament¹⁵.

Solar X-ray jets are imaged by satellite-borne telescopes in space in the ~ 0.2 – 2.0 -keV range. They are dynamic (with upward velocities of around 200 km s^{-1}), long (about $5 \times 10^4 \text{ km}$), narrow ($8 \times 10^3 \text{ km}$), and transient (with lifetimes of about 10 minutes)^{16,17}. In the accepted version of the emerging-flux model of jet formation^{3,18–21}, an emerging magnetic bipole enters a dominant-polarity (say, negative) ambient open magnetic field (that is, a field that extends far into the heliosphere), and the bipole’s minority-polarity (positive) side can reconnect with the coronal open field at the location of the magnetic-null region between the bipole and the ambient field. In this model, a burst of reconnection connects the outside of the bipole with the adjacent coronal field, producing a small loop on the outside of the emerging bipole’s minority-polarity foot, and reconnects the open field to the outside of the bipole’s majority-polarity foot. An X-ray jet develops as reconnection-heated material flows out along the new open-field strands. Moreover, in this model the presence of the X-ray-jet bright point (JBP) at the edge of the jet’s base is explained by the existence of the small loop that is formed by reconnection at the emerging field’s edge. In an extension of the emerging-flux model, the emerged bipole explodes as it reconnects, forming a ‘blowout jet’ with a relatively broad spire¹³. (See Methods and Extended Data Fig. 1 for further details of the emerging-flux model.)

To assess observationally the production of X-ray jets, we analysed 20 jets (Extended Data Table 1) in the solar polar regions using X-ray images from the X-ray telescope (XRT) on the Hinode satellite²²; this telescope detects a broad temperature range of coronal plasmas hotter than about 1.5 MK. We also used concurrent extreme ultraviolet (EUV) images from the Solar Dynamics Observatory’s (SDO’s)

Atmospheric and Imaging Assembly (AIA)²³, whose various filters detect plasmas primarily over narrow temperature ranges centred at, for example, approximately 0.05 MK, 0.6 MK, 1.6 MK or 2.0 MK, respectively, for wavelengths of 304 Å, 171 Å, 193 Å and 211 Å (see Methods).

Figure 1 shows a typical example of our results in both soft X-ray (Fig. 1a–c) and EUV (Fig. 1d–f) images. Between Fig. 1a and Fig. 1b, the jet’s spire, arched base, and JBP all begin brightening. Later (Fig. 1c), the spire extends higher, with the JBP positioned about $10''$ west of the spire. From a movie constructed from the XRT images (see Supplementary Video 1), we can see that the JBP starts to brighten at about 22:07 universal time (UT), with the spire becoming visible about 2.5 minutes later. Thus one could assume that the emergence of this jet fits with the emerging-flux model, whereby external reconnection (that is, reconnection occurring on the outside of the closed driving field²⁴) of the emerging field forms the JBP and gives rise to the spire at a displaced location. However, observing the same feature in AIA 193-Å EUV images (Fig. 1 and Supplementary Video 1) does not support this interpretation. These images clearly show a dark feature, similar to a small-scale solar chromospheric filament (hereafter ‘minifilament’), moving upwards and laterally, starting at around 22:06 UT. Its velocity is $\sim 40 \text{ km s}^{-1}$ between 22:07 UT and about 22:10 UT, when it reaches the apex of the illuminated arched base of the X-ray jet. After 22:10 UT, the minifilament is expelled in the spire of an EUV jet that is the counterpart to the XRT jet. In the EUV images, the jet has both emission and absorption components, with the minifilament evolving into part of the jet. Notably, however, in both soft X-ray and EUV images, the JBP is at the location from which the minifilament erupted. Thus the JBP is the analogue of the commonly observed solar flare arcade that forms in the wake of larger-scale filament eruptions; such flare arcades are made by internal reconnection (that is, reconnection occurring on the inside of the closed driving field²⁴) of the legs of the erupting closed field of a filament. This is not consistent with the JBP resulting from external reconnection, as proposed in the emerging-flux model.

We found an erupting minifilament to be discernible in AIA images of all 20 of the jets, with the minifilament’s eruption starting near the location of the JBP. In most cases, we could see that the JBP occurred where the minifilament (or part of the minifilament) had been rooted in the surface before ejection; we could not verify this arrangement in a few cases, in which the minifilament and JBP were along the same line of sight, but even then the observations are consistent with the JBP occurring at the location from which the minifilament was ejected. Typically, first the minifilament starts to lift off from the surface, and then the JBP starts to brighten. This is similar to the situation with large-scale filament eruptions, where the start of the eruption precedes the flare-brightening onset²⁵. Apart from their size, the eruptions of minifilaments in the production of X-ray jets are indistinguishable from the commonly observed eruptions of larger filaments in the onsets of solar flares. In some cases (see Extended Data Table 1, event 4, event 9 and event 13, and possibly event 1), rather than the entire minifilament lifting off, there is a whipping-like motion, with the

¹NASA/Marshall Space Flight Center, Huntsville, Alabama 35812, USA. ²Center for Space Plasma and Aeronomic Research, University of Alabama in Huntsville, Huntsville, Alabama 35899, USA.

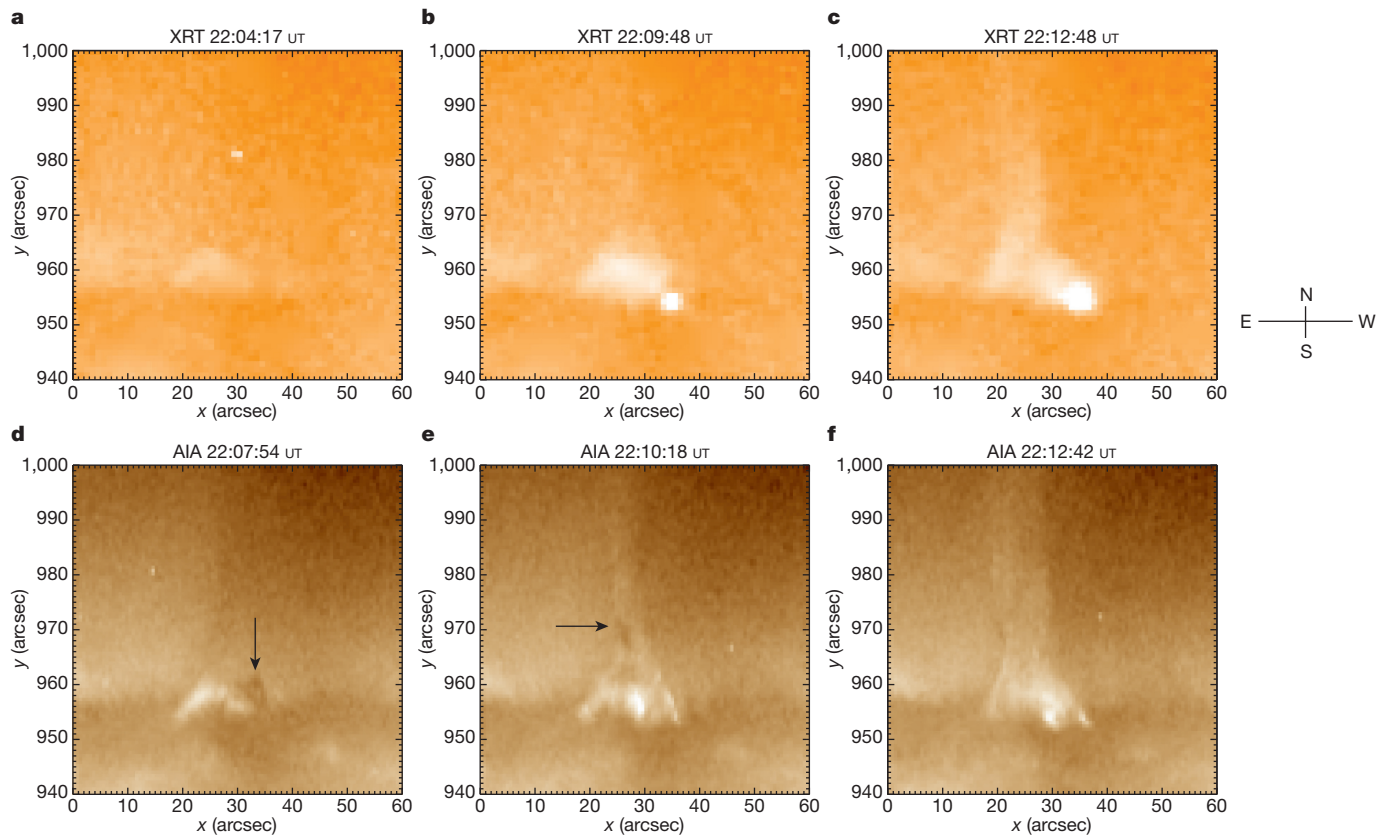


Figure 1 | Erupting-jet example. An example jet from 17 September 2010, as detected in soft X-ray (Hinode/XRT, TiPoly filter; **a–c**) and EUV (SDO/AIA 193 Å; **d–f**). In **b**, the jet bright point (JBP) is visible as a localized brightening; in **c**, the jet is fully developed and offset eastward of the JBP.

Arrows show a minifilament moving outwards from the JBP location. Panels **a** and **d** are 217 seconds apart; **b** and **e** are 30 seconds apart; and **c** and **f** are 6 seconds apart. See Methods for details, and Supplementary Video 1 for animations. This is event 18 of Extended Data Table 1.

JBP (flare) occurring below the whipping minifilament or at the location where the fastest moving part of the minifilament first detaches from the solar surface. Thus all cases are consistent with the JBP being a small flare arcade forming in the wake of the erupting minifilament^{4–7}.

We measured the lengths and velocities (as seen projected against the plane of the sky) of the minifilaments, during the period after they started to erupt but before they reached the jet-spire location. The average length of the minifilaments was $11''$ (8×10^3 km) with a standard deviation of $4''$. This is much smaller than the sizes quoted for filaments from an extensive survey²⁶ (3×10^4 km to 1.1×10^5 km), justifying the use of the term ‘minifilaments’. (Perhaps identical minifilaments had been previously identified on the solar disk²⁷.) Our measured average minifilament length is equal to the average width of X-ray jets¹⁷, consistent with the idea that the jet eruption is being driven by the minifilament eruption. We obtained mean velocities and a standard deviation for the erupting minifilaments of 31 ± 15 km s^{–1}. In all cases, the true sizes and speeds should tend to be larger than these plane-of-sky values.

X-ray jets have been classified as ‘standard’ or ‘blowout’ on the basis of the morphology of the spire and the intensity of the rest of the jet’s base compared with the JBP intensity: a standard jet has a narrow spire with a relatively dim base, while a blowout jet has a broad spire and a base that becomes about as bright as the JBP¹³. The emerging-flux model suggests that the difference occurs depending on whether the emerging-flux structure remains largely inert (standard jet), or erupts as the jet forms (blowout jet). Our new view is different. In a previous study¹³ of our 20 events, we morphologically classified 14 as blowout, 5 as standard, and 1 as ambiguous. We now find, however, that all 20 events seem to form in the same way—from erupting minifilaments. A jet has blowout-jet morphology if the erupting minifilament

strongly ejects from the base region (corresponding to an ejective larger-scale solar eruption⁶). Standard-jet morphology seems to result when the erupting minifilament mainly does not escape the closed-field base (maybe corresponding to confined larger-scale filament eruptions⁶), or perhaps if the eruption is ejective but very weak. We imagine that there is a continuum of morphological jet types, probably depending on the eruption’s strength and whether the erupting filament escapes the base.

From our observations we infer the schematic picture of Fig. 2 for jet production. Initially (Fig. 2a), two bipoles exist side by side, the larger one corresponding to what we usually observe as the base of the jet (compare with Fig. 1). The smaller bipole contains substantial free energy in the form of sheared and twisted magnetic field; that field holds a minifilament. As with the case of large-scale solar eruptions, this field becomes unstable by some process; it then erupts outwards, guided between the large bipole and the ambient open field. After the minifilament’s lift-off, internal reconnection occurs among the distended legs inside the minifilament field (Fig. 1b), making a ‘flare-arcade’ JBP. The spire starts to form as soon as the outer envelope of the minifilament-carrying erupting field begins external reconnection with the open field on the far side of the large bipole. External reconnection continues and soon reconnects the field threading the erupting minifilament with the far-side open field, injecting minifilament plasma along that open field. The external reconnection also adds a new hot layer to the larger bipole (larger red loop in Fig. 2c).

If the erupting minifilament-carrying field blows out beyond the large bipole’s apex (Fig. 2b, c), then widespread external reconnection results; this creates a broad jet spire characteristic of blowout jets. If the erupting field stalls near the apex of the large bipole (and/or if the eruption is weak enough), the external reconnection produces only a

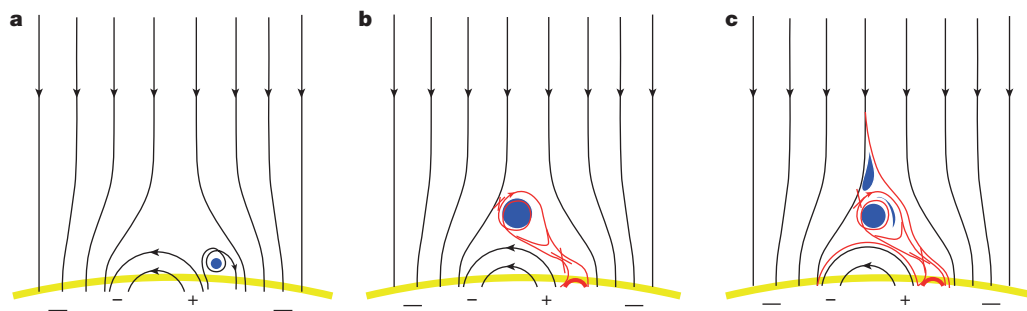


Figure 2 | Revised jet-eruption picture. Representation of the minifilament-eruption process that drives the formation of solar X-ray jets, as inferred from our observations. Black lines represent magnetic field, with arrows indicating polarities; red curves are newly reconnected field lines; blue features are minifilament material; yellow curve is the solar limb (the apparent edge of the

Sun). From the initial state (a), the jet forms as the minifilament erupts (b, c), with reconnection locations indicated by red crosses (b, c). The JBP (bold red arc) forms at the location of filament lift-off (b, c). See Methods for more details.

narrow jet, characteristic of a standard jet. Examples of blowout jets are shown in Fig. 1, and in Extended Data Figs 2 and 3 and their corresponding videos (Supplementary Videos 2 and 3). Examples of standard jets are shown in Extended Data Figs 4 and 5 and their corresponding videos (Supplementary Videos 4 and 5).

The emerging-flux model fails to explain our observation of a JBP occurring below the erupting minifilament, which the scenario shown in Fig. 2 naturally explains. Also, an expectation of the emerging-flux model is that, as the external reconnection progresses, reconnected open field will occur progressively closer to the JBP than does open field that reconnected earlier¹⁹. That is, the jet spire should drift towards the JBP in the emerging-flux model. Observations show, however, that more often than not the spire drifts away from the JBP²⁸. The schematic shown in Fig. 2 again explains this tendency for spire drift away from the JBP: the external reconnection of the erupting minifilament-carrying field produces reconnected open-field lines that in the corona stand progressively further away from the eruption's source location, which is the location of the internal-reconnection flare arcade that is the JBP.

We have not addressed what leads to the minifilament eruptions we have detected. Some recent studies of on-disk coronal jets found that the miniature filaments probably resulted from the cancellation of magnetic flux in the hours leading up to the eruption^{9,12,29,30}. We suspect however that, as with large-scale eruptions, various agents could trigger the eruption, including flux cancellation and flux emergence. In the latter case, the flux emergence would trigger the minifilament's eruption, rather than directly driving the jet as proposed in the emerging-flux model.

The minority-polarity flux in the base of an X-ray jet presumably arises from flux emergence of compact field loops into the dominant-polarity ambient field. It would therefore seem that many X-ray jets should be produced by these closed-field emergences, in the manner of the long-accepted emerging-flux model. However, that we found no X-ray jets formed in this way (at least for jets in polar coronal holes) suggests that external reconnection of the emerging closed field with the ambient open field occurs continuously and fast enough to keep an appreciable current sheet from building up at the magnetic-null region between the two fields, and that a burst of enough external reconnection to make an X-ray jet can be made only dynamically, driven by sudden eruption of the closed field as in a filament eruption. That is, the observed lack of X-ray jets formed in accordance with the emerging-flux model suggests that no current sheet of the scale of the overall system of two reconnecting fields can be formed gradually (quasi-stably) in the low-beta magnetized plasma of X-ray jets (where 'low-beta' refers to a ratio of gas-to-magnetic pressures of much less than one), and by analogy not in similar reconnection events in other low-beta astrophysical settings either.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 August 2014; accepted 7 May 2015.

Published online 6 July 2015.

- Shibata, K. *et al.* Observations of x-ray jets with the Yohkoh soft x-ray telescope. *Publ. Astron. Soc. Jpn.* **44**, 173L–179L (1992).
- Cirtain, J. W. *et al.* Evidence for Alfvén waves in solar jets. *Science* **318**, 1580–1582 (2007).
- Yokoyama, T. & Shibata, K. Magnetic reconnection as the origin of x-ray jets and H α surges on the sun. *Nature* **375**, 42–44 (1995).
- Hirayama, T. Theoretical model of flares and prominences. *Sol. Phys.* **34**, 323–338 (1974).
- Shibata, K. *et al.* Hot-plasma ejections associated with compact-loop solar flares. *Astrophys. J.* **451**, L83–L85 (1995).
- Moore, R. L., Sterling, A. C., Hudson, H. S. & Lemen, J. R. Onset of the magnetic explosion in solar flares and coronal mass ejections. *Astrophys. J.* **552**, 833–848 (2001).
- Chen, P. F. Coronal mass ejections: models and their observational basis. *Living Rev. Sol. Phys.* **8**, 1–92 (2011).
- Moore, R. L., Tang, F., Bohlin, J. D. & Golub, L. H-alpha macrospicules—identification with EUV macrospicules and with flares in x-ray bright points. *Astrophys. J.* **218**, 286–290 (1977).
- Hong, J. *et al.* Coronal bright points associated with minifilament eruptions. *Astrophys. J.* **796**, 73 (2014).
- Raouafi, N. -E., Georgoulis, M. K., Rust, D. M. & Bernasconi, P. N. Micro-sigmoids as progenitors of coronal jets: is eruptive activity self-similarly multi-scaled? *Astrophys. J.* **718**, 981–987 (2010).
- Nisticò, G., Bothmer, V., Patsourakos, S. & Zimbardo, G. Characteristics of EUV coronal jets observed with STEREO/SECCHI. *Sol. Phys.* **259**, 87–108 (2009).
- Adams, M., Sterling, A. C., Moore, R. L. & Gary, G. A. A small-scale eruption leading to a blowout macrospicule jet in an on-disk coronal hole. *Astrophys. J.* **783**, 11 (2014).
- Moore, R. L., Sterling, A. C., Falconer, D. A. & Robe, D. The cool component and the dichotomy, lateral expansion, and axial rotation of solar x-ray jets. *Astrophys. J.* **769**, 134 (2013).
- De Pontieu, B. *et al.* The origins of hot plasma in the solar corona. *Science* **331**, 55–58 (2011).
- Shibata, K. Evidence of magnetic reconnection in solar flares and a unified model of flares. *Astrophys. Space Sci.* **264**, 129–144 (1999).
- Shimojo, M. *et al.* Statistical study of solar x-ray jets observed with the Yohkoh soft x-ray telescope. *Publ. Astron. Soc. Jpn.* **48**, 123–136 (1996).
- Savcheva, A. *et al.* A study of polar jet parameters based on Hinode XRT observations. *Publ. Astron. Soc. Jpn.* **59**, S771–S778 (2007).
- Nishizuka, N. *et al.* Giant chromospheric anemone jet observed with Hinode and comparison with magnetohydrodynamic simulations: evidence of propagating Alfvén waves and magnetic reconnection. *Astrophys. J.* **683**, L83–L86 (2008).
- Moreno-Insertis, F. & Galsgaard, K. Plasma jets and eruptions in solar coronal holes: a three-dimensional flux emergence experiment. *Astrophys. J.* **771**, 20 (2013).
- Archontis, V. & Hood, A. W. A numerical model of standard to blowout jets. *Astrophys. J.* **769**, L21 (2013).
- Fang, F., Fan, Y. & McIntosh, S. W. Rotating solar jets in simulations of flux emergence with thermal conduction. *Astrophys. J.* **789**, L19 (2014).
- Kosugi, T. *et al.* The Hinode (solar-B) mission: an overview. *Sol. Phys.* **243**, 3–17 (2007).
- Lemen, J. R. *et al.* The atmospheric imaging assembly (AIA) on the solar dynamics observatory (SDO). *Sol. Phys.* **275**, 17–40 (2012).
- Sterling, A. C. & Moore, R. L. Internal and external reconnection in a series of homologous solar flares. *J. Geophys. Res.* **106**, 25227–25238 (2001).
- Sterling, A. C. & Moore, R. L. Slow-rise and fast-rise phases of an erupting solar filament, and flare emission onset. *Astrophys. J.* **630**, 1148–1159 (2005).

26. Bernasconi, P. N., Rust, D. D. M. & Hakim, D. Advanced automated solar filament detection and characterization code: description, performance, and results. *Sol. Phys.* **228**, 97–117 (2005).
27. Wang, J. *et al.* Minifilament eruption on the quiet sun. I. Observations at H α central line. *Astrophys. J.* **530**, 1071–1084 (2000).
28. Savcheva, A., Cirtain, J. W., DeLuca, E. E. & Golub, L. Does a polar coronal hole's flux emergence follow a Hale-like law? *Astrophys. J.* **702**, L32–L36 (2009).
29. Shen, Y., Liu, Y., Su, J. & Deng, Y. On a coronal blowout jet: the first observation of a simultaneously produced bubble-like CME and a jet-like CME in a solar event. *Astrophys. J.* **745**, 164 (2012).
30. Young, P. R. & Muglach, K. A. Solar dynamics observatory and Hinode observations of a blowout jet in a coronal hole. *Sol. Phys.* **289**, 3313–3329 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements A.C.S. and R.L.M. were supported by funding from the Heliophysics Division of NASA's Science Mission Directorate through the Living With A Star Targeted Research and Technology Program (LWS TR&T), and the Hinode Project. Both benefited from TR&T discussions and from discussions with S. K. Antiochos. We thank D. M. Zarro for assistance with video development. A.C.S. benefited from

discussions held at the International Space Science Institute (ISSI; Switzerland) International Team on Solar Coronal Jets (led by N. Raouafi). Hinode is a Japanese mission developed and launched by the Institute of Space and Astronautical Science (ISAS) of the Japan Aerospace Exploration Agency (JAXA), with the National Astronomical Observatory Japan (NAOJ) as a domestic partner, and NASA and the Science and Technology Facilities Council (UK) as international partners. It is operated by these agencies in cooperation with the European Space Agency and Norwegian Space Agency.

Author Contributions A.C.S. carried out the reduction, analysis, and interpretation of XRT and AIA data, software development, and manuscript preparation. R.L.M. interpreted the results and reviewed the manuscript. D.A.F. developed software, and assimilated and calibrated AIA data. M.A. discovered and analysed the seminal jet event that motivated this broader investigation, and carried out manuscript formatting and review.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.C.S. (alphonse.sterling@nasa.gov) or R.L.M. (ron.moore@nasa.gov).

METHODS

Emerging-flux model. According to the emerging-flux model^{3,18–21} (Extended Data Fig. 1) for the formation of solar coronal jets, an emerging bipole enters a dominant-polarity (negative in Extended Data Fig. 1) ambient open field, and the bipole's minority-polarity (positive) side can reconnect with coronal field at the location of the magnetic-null region between the bipole and the ambient field. After enough of the bipole has emerged, a burst of reconnection joins the outside of the bipole with the nearby coronal field (Extended Data Fig. 1b), resulting in two reconnection products: a small loop on the outside of the base of the emerging bipole's minority-polarity side, and an open field connecting the bipole's majority-polarity side with the open coronal field, giving a new footpoint connection for that coronal field. This type of reconnection has been called 'interchange'³¹, or 'external'²⁴, because the reconnection is on the outside of the closed driving field (the emerging field in this case). An X-ray jet develops as reconnection-heated material flows out along the new open-field strands. Additionally, the external-reconnection-formed small loop at the emerging field's edge is the model's explanation for the JBP (also called a 'hot loop'³) observed at the edge of the jet's base. According to the previous view of blowout jets¹³, the idea was that the external reconnection causes and/or is driven by ejective eruption (blowout) of the emerging bipole, which is assumed to contain substantial nonpotential (that is, twisted) magnetic field, driving that bipole's eruption along the ambient open field to make a broad jet spire¹³.

Instrumentation and data. For our X-ray images, we use data from the Hinode/XRT with 30-s cadence and 1" pixels. XRT detects a broad range of temperatures, but has highest sensitivity for temperatures of greater than about 1.5 MK, even with the TiPoly filter used for the observations presented here. (Among XRT's filters, the TiPoly filter detects relatively cool X-ray-emitting plasmas.) For each jet in Extended Data Table 1, we studied concurrent EUV images from SDO/AIA, which has 0.6" pixels and 12-s cadence. Our final movies and figures were formed by summing the frames in pairs, and therefore the resulting movies were generally of 1-min cadence and 24-s cadence respectively for XRT and AIA. This summing blurs the images somewhat, but renders subtle features, such as X-ray jets and some of the fainter EUV-detected minifilaments, much easier to discern. For many of the X-ray jets of our study, we examined all of the AIA EUV channels, which are tuned to wavelengths of 304 Å, 171 Å, 193 Å, 211 Å, 131 Å, 335 Å, and 94 Å; these have strong responses to logarithmic temperatures (in Kelvin) of about 4.7, 5.8, 6.2, 6.3, 7.0, 6.4, and 6.8 respectively (although some channels are multivalued²³). Usually there was little new information in the hotter 131-Å, 335-Å, and 94-Å channels, and so we did not inspect these hotter channels for some of the jets. We applied standard processing routines from the Solarsoft software library³² to the XRT and AIA images.

In total we examined 20 X-ray jets, initially selected during an earlier study¹³, in which the JBP was obvious in the X-ray images (Extended Data Table 1). Each event of Extended Data Table 1 is categorized as 'standard', 'blowout', or 'ambiguous' on the basis of its morphology in the XRT images (and, in some cases, in the AIA 304-Å images as well). Blowout jets are those in which the entire base brightened and the spire broadened to span approximately the width of the base; standard jets are those in which only the JBP brightened substantially in the base and the spire remained narrow compared with the span of the base. (The JBP is also referred to as the 'hot loop'³, 'bright loop'³, 'bright point'^{31,33}, and 'bright footpoint'¹⁷.)

In each blowout jet in Extended Data Table 1, the minifilament eruption seemed to be ejective; the erupting closed field apparently blows out into the ambient open field. In this case, much or all of the filament material escapes from the closed field onto the open field.

In the events of Extended Data Table 1 that are categorized as standard jets, a minifilament eruption was detectable, but usually that eruption either did not seem to be ejective, or was perhaps ejective but weak or faint. In event 4, a minifilament (best seen at 304 Å) has a whipping motion from the location that becomes the JBP. Event 7 seems to be generated by a minifilament that becomes partially destabilized and spins (rolls) beneath confining magnetic fields. These standard-jet events may therefore be analogous to larger-scale confined filament eruptions, ones that make flares that are of shorter duration than the ejective flares⁶. (As an example, the rolling minifilament of event 7 could be a scaled-down version of the confined filament eruption shown in figure 1 of ref. 34, and in the corresponding online movies of ref. 34.) However, event 6—another standard jet—shows an ejective minifilament, similar to the jets identified as blowout jets, but it does not make a broad spire. In that case it appears that the minifilament erupted far enough for much of it to escape into the open field through external reconnection, but not enough to blow out violently and form a broad jet. In comparison with the blowout jets, more of the filament material remains trapped within the closed field.

Our other standard jets (events 5 and 19), and the ambiguous jet (event 11), may also be partially confined and partially ejective minifilament eruptions. In these cases, some of the minifilament material escapes onto the open field, and some of it remains in the closed field. In this sense, we envisage a continuum of jet manifestations, between pure blowout jet (where the filament field would push far into the opposite-polarity open field, making a broad jet, and all of the filament material would eventually escape onto that open field), and a pure standard jet (where only the envelope of the closed filament field reconnects with the opposite-polarity open field, and none of the closed field containing the cool filament material undergoes external reconnection). Our view of standard jets as being due to confined minifilament eruptions, partially confined minifilament eruptions, and/or weak ejective minifilament eruptions is still speculative. Further study will be required to understand fully the various morphological differences among jets.

Minifilament measurement details. We measured the lengths and velocities (projected in the plane of the sky normal to the Earth–Sun line-of-sight) of the minifilaments during the period after they started to erupt but before they formed a jet or reached the apex of the base (below the jet spire). We usually used the 171-Å, 193-Å or 211-Å AIA channels for these measurements; only for events 4, 7 and 10 did we find the 304-Å channel preferable for determining minifilament properties in our data set. We obtained mean velocities for the erupting minifilaments of $31 \pm 15 \text{ km s}^{-1}$; if the velocities are weighted inversely with their uncertainty (Extended Data Table 1), the weighted mean velocity and weighted standard deviation are 24 km s^{-1} and 13 km s^{-1} , respectively.

The jet-formation process in our picture. As shown in Fig. 2, we envisage that initially a minifilament-carrying, nonpotential, relatively compact core field of a magnetic bipole (or magnetic arcade) exists next to (and shares the minority-polarity flux with) a relatively large bipole (Fig. 2a). An unspecified process destabilizes the smaller bipole so that it erupts, with the minifilament being channelled between the large bipole and the overlying open field. Upon reaching the open coronal field on the far side of the large bipole, the field carrying the minifilament reconnects with that field (Fig. 2b), and a jet, often including substantial minifilament material, is ejected along the newly reconnected open field (Fig. 2c). This reconnection also adds field lines to the large bipole. Internal reconnection (the lower red cross) of the minifilament-carrying field also occurs (Fig. 2b); this reconnection is inside the erupting lobe of the double bipole, and forms a flare arcade (the JBP) in the wake of the ejected minifilament.

This process of X-ray-jet formation is analogous to the formation of commonly observed flare arcades in typical large-scale solar eruptions; that is, the erupting lobe of the system erupts as in a 'typical' large-scale eruption, as pictured in, for example, figure 1 of ref. 5 or figure 1 of ref. 6. In our jet-formation picture, this process is occurring on a smaller scale, so that the filament of those typical models corresponds to our minifilaments. However, rather than a filament travelling directly outwards as in those large-scale eruptions, in the case of X-ray jets the minifilament travels along the curved path between the adjacent bipole and distorted ambient coronal field. (The coronal field is distorted by the magnetic field of the two bipoles.) As long as the erupting minifilament is on the near side (that is, the side of its origin) of the apex of the neighbouring bipole, no reconnection occurs between the erupting-bipole field enveloping the filament and the ambient coronal fields. (In three dimensions the situation will not be as simple as in the two-dimensional schematic, but we still expect the basic picture to hold.) We will consider what happens when the enveloping field reaches the far side of the apex shortly. First, however, looking again at the schematics of the typical large-scale eruptions^{5,6}, it can be seen that the field lines beneath the erupting filament reconnect (this is what we are calling internal reconnection) to form hot flare loops near the solar surface. In our analogous schematic (Fig. 2), these flare loops correspond to the JBP. While the small lobe of the double bipole in Fig. 2 is erupting in this fashion, the neighbouring bipole remains largely inert, except for the addition of the new field via external reconnection, as mentioned above.

We now consider what happens when the erupting-minifilament bipole reaches the far side of the apex of the neighbouring bipole (Fig. 2b). Because the field orientations are then opposite, the erupting field enveloping the minifilament and the far-side ambient coronal field can undergo reconnection; because this reconnection is between the field of the erupting bipole and the coronal field that is external to that erupting bipole, we call this external reconnection. This external reconnection adds heat to the reconnected field lines, making a hot jet spire along the open field lines and forming hot loops over the adjacent bipole (red curves in Fig. 2c). This external reconnection progressively erodes the field enveloping the cool minifilament material. If this erosion of the enveloping field stops before the field lines holding cool material is reached—which could happen if, for example, the erupting minifilament-carrying bipole does not have enough energy to travel deep into the far-side ambient-field region—then the cool material never reaches the open field (and the spire receives no cool material). Rather, the filament plasma remains trapped in the closed field in the base of the jet. This may be how the

standard jets are formed; only a narrow hot spine forms if the erupting minifilament-carrying bipole does not go far into the ambient-field region.

In a blowout jet, the eruption continues deeper into the ambient field region of oppositely directed polarity to make a broader spine than is depicted in Fig. 2c. The envelope around the cool-minifilament material is completely eroded away, and so the cool material escapes onto the open ambient coronal field, forming a cool jet. In this sense, the eruption of the minifilament is analogous to ejective eruptions of typical large-scale cases. (Some standard jets appear to be weak versions of such ejective jets.) The drawings in Fig. 2 are tailored to depict the jet in Fig. 1 (jet 18 in Extended Data Table 1), which is a blowout jet.

The external reconnection of the erupting-minifilament field with the open field also adds a new hot layer to the larger bipole (larger red loop in Fig. 2c); this reconnection product from earlier eruption episodes might have created the 'initial' large bipole (large black loops of Fig. 2a). Other possibilities for the initial large bipole are that it and the filament-carrying bipole are two asymmetric lobes of a so-called anemone field region³⁵. That anemone region could be due to recently emerged magnetic flux, or it could have formed over time via surface-flux migration and cancellation³⁶.

A schematic for X-ray jets similar to that of our Fig. 2 is shown in figure 8b of ref. 15. That figure was derived from data from earlier satellite missions, before the high-resolution, high-cadence, multiple-EUV-wavelength data of SDO/AIA. There is, however, a difference between the picture of ref. 15 and our picture. The proposal there is that a plasmoid (which might correspond to our minifilament) erupts from the external-reconnection site of the emerging-flux model (Extended Data Fig. 1), the pre-eruption plasmoid being in the current sheet between the emerging flux and the ambient coronal field. (Also, figure 6 of ref. 15 explicitly depicts an emerging-flux origin for X-ray jets.) In contrast, our proposal is that X-ray jets, at least in coronal holes, are a miniature version of large-scale flares and coronal mass ejections, regardless of whether there is emerging flux. In our view, before eruption the minifilament resides in sheared field (or in a twisted-field magnetic flux rope) in the core of a magnetic arcade, instead of in a current sheet. More generally, in our view the triggering and eruption of the minifilament may include any of a multitude of processes and subprocesses proposed for large-scale eruptions, including those listed in the main text, and others^{7,37–39}. Determining whether the pre-eruption minifilaments that erupt in jets are located at an external-reconnection current sheet (as suggested in ref. 15), or instead reside in a magnetic arcade, as we envisage, requires further observational study.

In our AIA movies the developing jets show clear rotation in some cases, such as the jet of Fig. 1 (Supplementary Video 1). Other jets, however, show only partial rotation (for example, Extended Data Fig. 2 and Supplementary Video 2), or no obvious rotation (for example, Extended Data Fig. 3 and Supplementary Video 3). Because we have not identified a clear pattern regarding the rotations and the resulting jets, we do not address this topic further here.

Cause of minifilament-eruption onset. Given that we have not examined jets that originate at low solar latitudes, we cannot adequately see the causes (triggering) of these magnetic eruptions. As with large-scale filament eruptions, several triggering agents could be responsible, including flux cancellation or even flux emergence. Our main point here is that, independently of the cause of the minifilament-eruption onset, the jets all result from those minifilament eruptions, with the JBP being the 'flare' that occurs in conjunction with those minifilament eruptions.

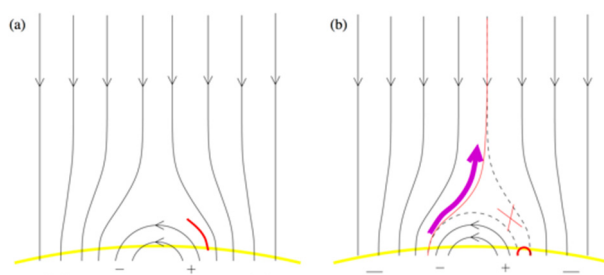
As stated in the main text, however, several other studies^{9,12,29,30} found on-disk coronal jets to occur in conjunction with magnetic flux cancellation. One study¹²

searched for emerging flux beneath a jet, but found no noticeable signature of emergence. A different study⁴⁰ also searched for but did not find emerging flux below an on-disk coronal jet. Another study⁴¹ found mini coronal mass ejections, perhaps resulting from 'small filament ejections', that may be similar or identical to the jets we discuss here; that study reports the ejections to occur at sites of "twisting small concentrations of opposite polarity magnetic field", and again there was no detection of emerging flux. Similar jets have been reported elsewhere⁴², but without direct magnetic field observations.

We have found two studies of on-disk jets where emerging flux was reported. In the first⁴³, although emergence occurred, a microflare and an EUV jet happened only after cancellation of flux in the region of the flux emergence. Similarly, in the second study⁴⁴ flux emergence occurred, but two jets occurred at about the time that the emerged flux underwent cancellation with the neighbouring field. In that case⁴⁴, the jet observations were from XRT, and were of jets occurring in on-disk coronal holes; so those observations are on-disk complementary examples of the near-limb XRT polar-coronal-hole jet observations that we present here.

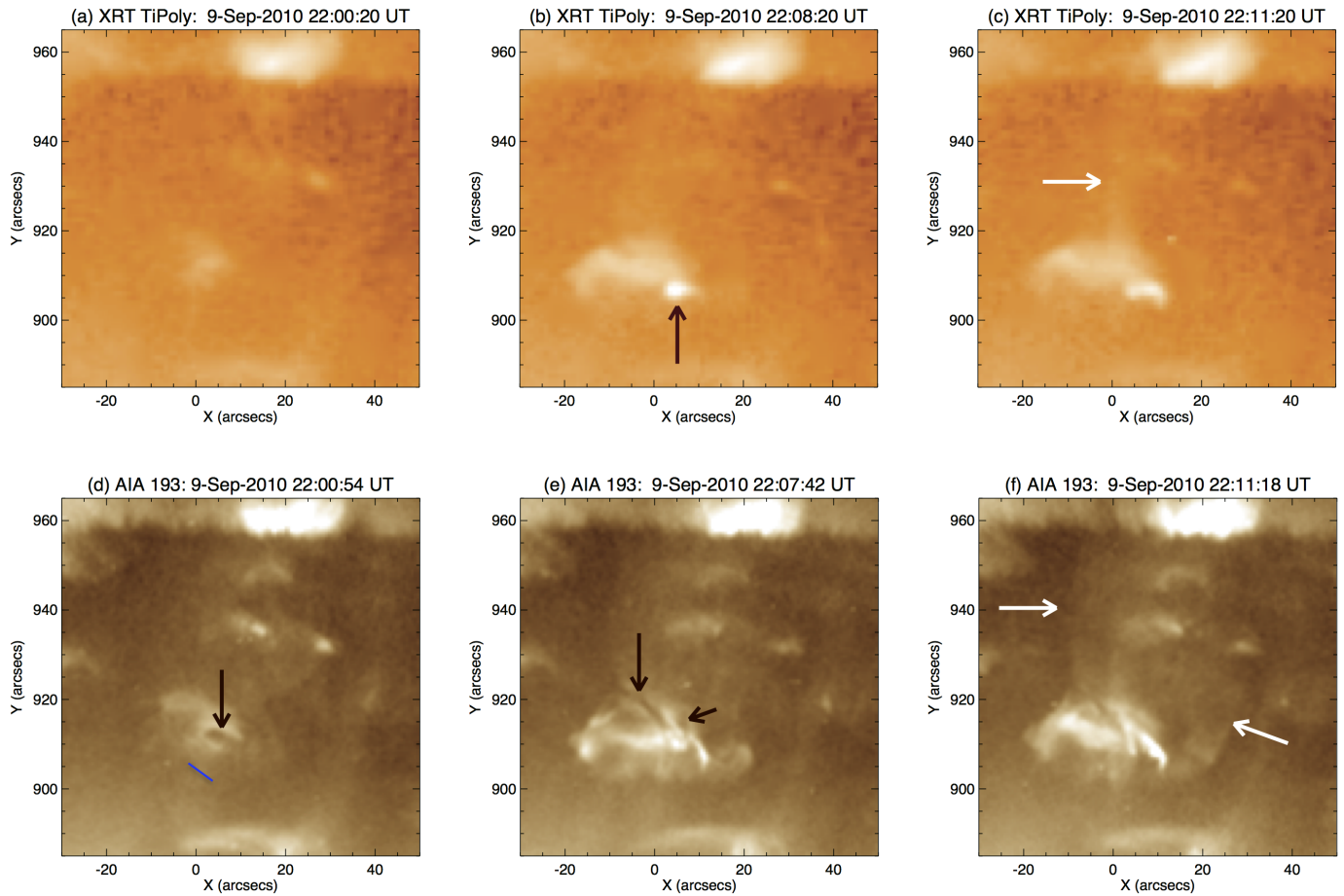
On balance, then, the on-disk coronal jet studies suggest that flux cancellation is often crucial to jet onset. In light of our findings, we expect that, in those earlier observations, the cancellation probably resulted in minifilament eruptions that produced jets, with flares occurring in the wake of those eruptions and appearing as JBPs.

31. Crooker, N. U., Gosling, J. T. & Kahler, S. W. Reducing heliospheric magnetic flux from coronal mass ejections without disconnection. *J. Geophys. Res.* **107**, 1028–1032 (2002).
32. Freeland, S. L. & Handy, B. N. Data analysis with the SolarSoft system. *Sol. Phys.* **182**, 497–500 (1998).
33. Moore, R. L., Cirtain, J. W., Sterling, A. C. & Falconer, D. A. Dichotomy of solar coronal jets: standard jets and blowout jets. *Astrophys. J.* **720**, 757–770 (2010).
34. Sterling, A. C., Moore, R. L. & Freeland, S. L. Insights into filament eruption onset from Solar Dynamics Observatory observations. *Astrophys. J.* **731**, L3 (2011).
35. Shibata, K. *et al.* Chromospheric anemone jets as evidence of ubiquitous reconnection. *Science* **318**, 1591–1594 (2007).
36. Innes, D. & Teriaca, L. Quiet sun explosive events: jets, splashes, and eruptions. *Sol. Phys.* **282**, 453–469 (2013).
37. Shibata, K. & Tanuma, S. Plasmoid-induced-reconnection and fractal reconnection. *Earth Planets Space* **53**, 473–482 (2001).
38. Lin, J., Forbes, T. G. & Isenberg, P. A. Prominence eruptions and coronal mass ejections triggered by newly emerging flux. *J. Geophys. Res.* **106**, 25053–25074 (2001).
39. Kusano, K. *et al.* Magnetic field structures triggering solar flares and coronal mass ejections. *Astrophys. J.* **760**, 31 (2012).
40. Chandrashekhara, K., Morton, R. J., Banerjee, D. & Gupta, G. R. The dynamical behaviour of a jet in an on-disk coronal hole observed with AIA/SDO. *Astron. Astrophys.* **562**, A98 (2014).
41. Innes, D. E., Genetelli, A., Atie, R. & Potts, H. E. Quiet Sun mini-coronal mass ejections activated by supergranular flows. *Astron. Astrophys.* **495**, 319–323 (2009).
42. Schrijver, C. J. Eruptions from solar ephemeral regions as an extension of the size distribution of coronal mass ejections. *Astrophys. J.* **710**, 1480–1485 (2010).
43. Liu, C., Deng, N., Liu, R., Ugarte-Urra, I., Wang, S. & Wang, H. A standard-to-blowout jet. *Astrophys. J.* **735**, L18 (2011).
44. Huang, Z., Madjarska, M. S., Doyle, J. G. & Lamb, D. A. Coronal hole boundaries at small scales. IV. SOT view. Magnetic field properties of small-scale transient brightenings in coronal holes. *Astron. Astrophys.* **548**, A62–A80 (2012).
- 45.ariat, E., Antiochos, S. K. & De Vore, C. R. A model for solar polar jets. *Astrophys. J.* **74**, 61–74 (2009).



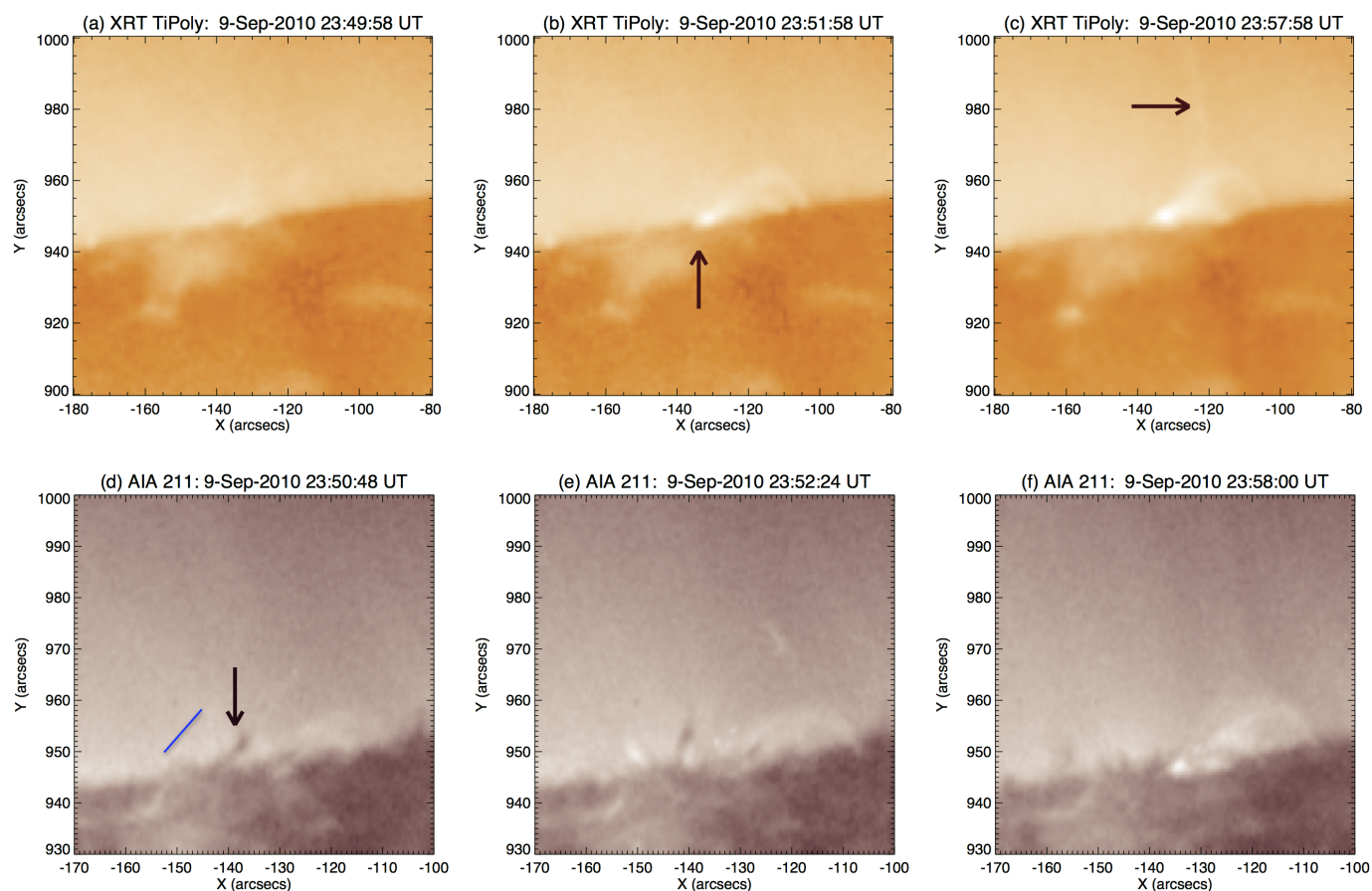
Extended Data Figure 1 | Emerging-flux model for the formation of solar X-ray jets. The commonly accepted mechanism for jet formation¹. Black lines represent magnetic field, with arrows indicating polarity; the yellow curve is the solar limb; the thick red curve in **a** represents a plasma current sheet; the red cross in **b** shows the location of field reconnection. **a**, Initial state. **b**, Jet

formation: flux emergence purportedly forces reconnection at the current sheet (red cross), resulting in new closed-loop field (red loop), and new connections to the open coronal field (thin red line), along which the X-ray jet (purple) flows. According to this model, the new reconnection loops appear as the JBP. Previous scenarios for 'blowout jets'^{213,33,45} have been variations of this model.



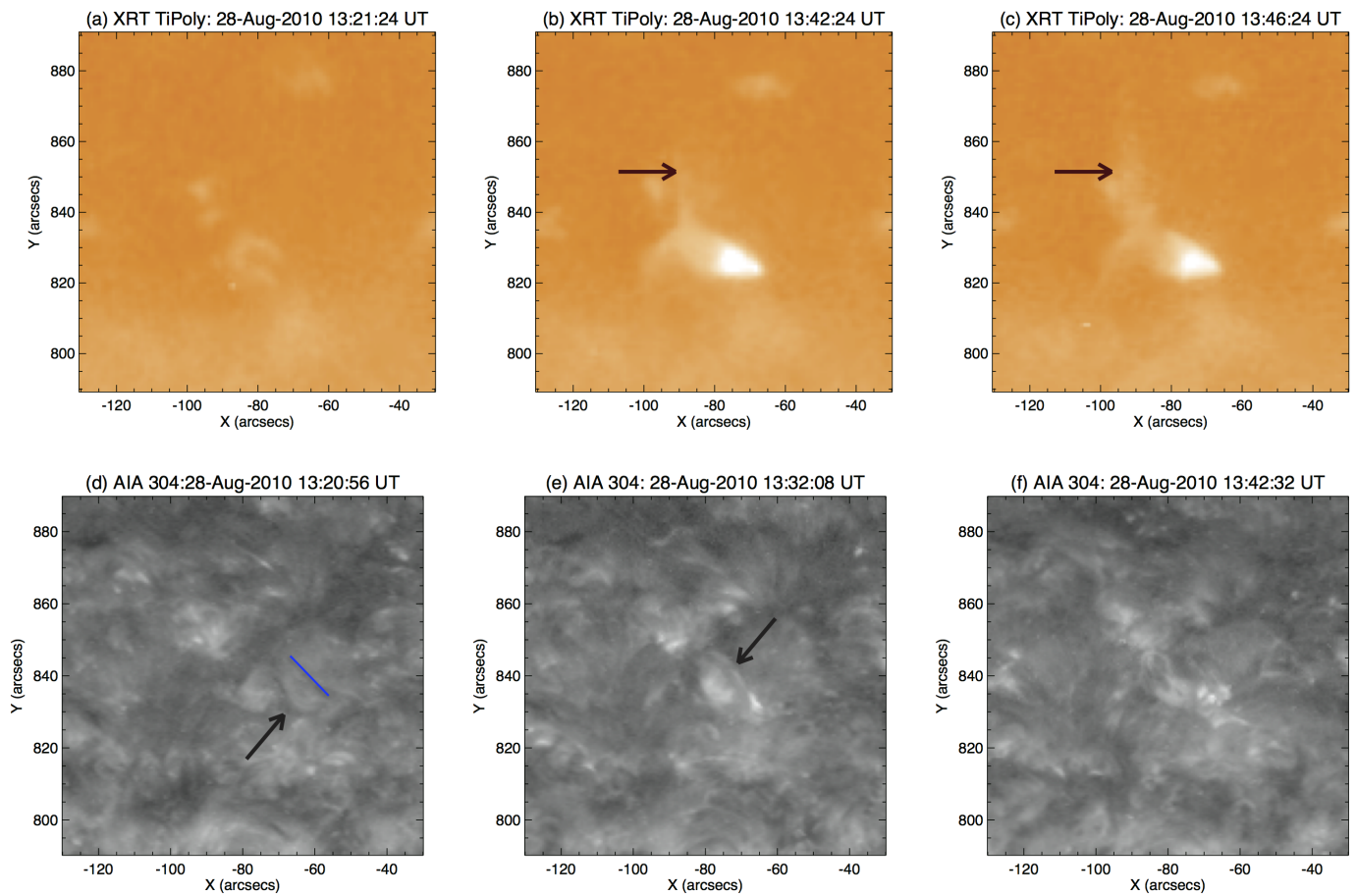
Extended Data Figure 2 | Jet of 2010 September 9, 22 UT. **a–c**, XRT, and **d–f**, 193-Å AIA images of the jet. Arrows show: **b**, the developing JBP; **c**, the X-ray-jet spire; and **d**, the minifilament. In **e**, both arrows point to segments of the minifilament, which split during eruption; in **f**, both arrows point to the edges of a broad jet. In **d**, the blue bar shows our estimate of the size of the

minifilament, the value of which appears in Extended Data Table 1. See Supplementary Video 2 for animations. This is event 12 of Extended Data Table 1. North is to the top and west to the right of these images (and all other solar images in this paper).



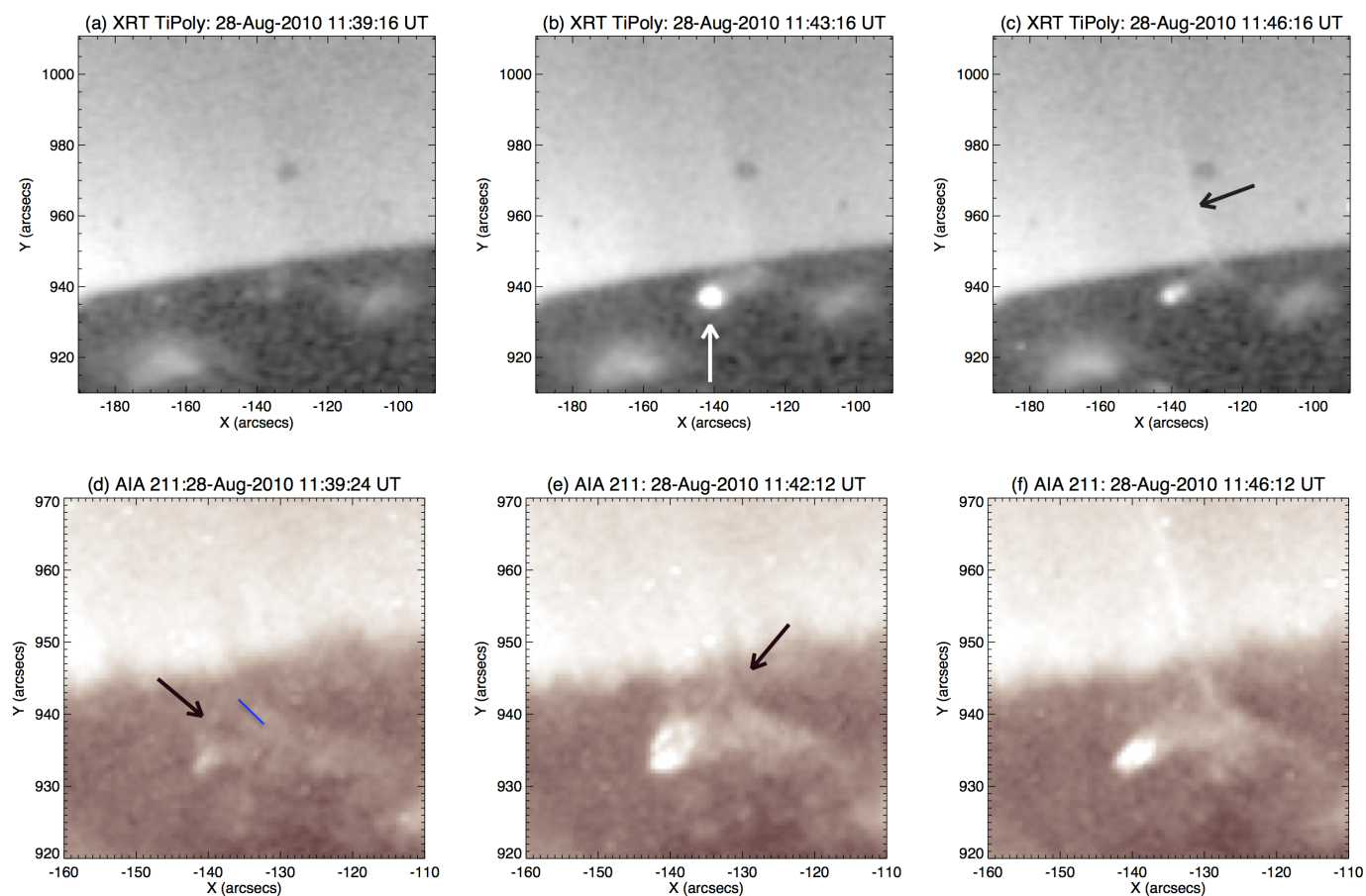
Extended Data Figure 3 | Jet of 2010 September 9, 23 UT. **a–c**, XRT, and **d–f**, 211-Å AIA images of the jet. Arrows show: **b**, the developing JBP; **c**, the X-ray-jet spire; and **d**, the minifilament starting to erupt. The blue bar in **d**

shows our estimate of the size of the minifilament. The AIA images show a smaller field of view than the XRT images. See Supplementary Video 3 for animations. This is event 13 of Extended Data Table 1.



Extended Data Figure 4 | Jet of 2010 August 28, 13 UT. **a–c**, XRT, and **d–f**, 304-Å AIA images of a ‘standard’ jet. Arrows show: **b**, the X-ray jet spire; **c**, the X-ray jet spire, showing drift since **b**; **d**, the minifilament starting to erupt; **e**, ‘rolling’ filament (see Methods). The blue bar in **d** shows our estimate

of the size of the minifilament. The grey-scale images show the filament better than the colour images for this event. See Supplementary Video 4 for animations. This is event 7 of Extended Data Table 1.



Extended Data Figure 5 | Jet of 2010 August 28, 11 UT. **a–c**, XRT, and **d–f**, 211-Å AIA images of a ‘standard’ jet. The dark spot northwest of centre in the XRT images is an artefact. Arrows show: **b**, the JBP; **c**, the X-ray jet spire; **d**, the minifilament moving upwards; **e**, the minifilament near the apex of

the jet base, with the jet spire starting to develop. The AIA images show a smaller field of view than the XRT images. The blue bar in **d** shows our estimate of the size of the minifilament. See Supplementary Video 5 for animations. This is event 6 of Extended Data Table 1.

Extended Data Table 1 | The X-ray jets studied here

Event	Date ^a	Start; End ^b	x, y (arcsec) ^c	Type ^d	Fil. Size ^e (arcsec)	Fil. Speed ^e (km s ⁻¹)
1	2010 Jul 24	15:56; >16:15	-60, 950	blowout	17	14 ± 2
2	2010 Jul 25	12:29; 12:46	140, -950	blowout	10	30 ± 10
3	2010 Aug 26	14:13; >14:16	100, 950	blowout	10	28 ± 5
4	2010 Aug 27	11:35; 12:17	30, 920	standard	20 ^f	50 ± 5 ^f
5	2010 Aug 27	11:40; 12:20	-50, 920	standard	diffuse(?) ^f	28 ± 5(?) ^f
6	2010 Aug 28	11:40; 12:03	-130, 940	standard	5	28 ± 5
7	2010 Aug 28	<13:41; >13:48	-70, 840	standard	17	rolling
8	2010 Sep 05	21:14; 21:35	30, 840	blowout	10	28 ± 5
9	2010 Sep 08	01:29; 01:44	40, 935	blowout	6	19 ± 5
10	2010 Sep 09	20:14; 20:33	20, 770	blowout	17	73 ± 8
11	2010 Sep 09	20:21; 20:40	60, 850	ambiguous	12	uncertain ^g
12	2010 Sep 09	22:05; 22:31	0, 910	blowout	7	13 ± 3
13	2010 Sep 09	23:52; 00:06	-120, 950	blowout	9	33 ± 5
14	2010 Sep 10	00:01; 00:09	-10, 880	blowout	7	50 ± 10
15	2010 Sep 11	00:39; 00:50	80, 950	blowout	8 ^f	19 ± 5 ^f
16	2010 Sep 11	<01:08; 01:27	-120, 950	blowout	13	40 ± 8
17	2010 Sep 17	20:39; 21:08	-20, 840	blowout	diffuse ^h	33 ± 8 ^h
18	2010 Sep 17	22:08; 22:18	30, 960	blowout	7	40 ± 5
19	2010 Sep 19	19:47; 20:23	20, 880	standard	10	20 ± 5
20	2010 Sep 27	00:39; 00:43	0, 960	blowout	10	20 ± 5

^aDate the event started. ^bTime period (UT) during which a clearly detectable jet and/or compact JBP is visible in XRT images. < and > indicate that the jet started before or continued after, respectively, the indicated times during gaps in XRT data. ^cApproximate x, y location of the jet in AIA images in heliocentric coordinates. ^dMorphological classification of the X-ray jet based on ref. 13. ^eLine-of-sight projected size/speed of the minifilament near the time of eruption onset; size uncertainty less than about 3". ^fMinifilament diffuse or faint, or identification less certain than in other cases. ^gAccurate speed measurement not possible owing to image shifts during eruption time. ^hMinifilament too diffuse for size measurement, but moving structures can be tracked for velocity estimate.

DNA rendering of polyhedral meshes at the nanoscale

Erik Benson^{1,2}, Abdulmelik Mohammed³, Johan Gardell^{1,2}, Sergej Masich⁴, Eugen Czeizler³, Pekka Orponen³ & Björn Högberg^{1,2}

It was suggested¹ more than thirty years ago that Watson–Crick base pairing might be used for the rational design of nanometre-scale structures from nucleic acids. Since then, and especially since the introduction of the origami technique², DNA nanotechnology has enabled increasingly more complex structures^{3–18}. But although general approaches for creating DNA origami polygonal meshes and design software are available^{14,16,17,19–21}, there are still important constraints arising from DNA geometry and sense/antisense pairing, necessitating some manual adjustment during the design process. Here we present a general method of folding arbitrary polygonal digital meshes in DNA that readily produces structures that would be very difficult to realize using previous approaches. The design process is highly automated, using a routing algorithm based on graph theory and a relaxation simulation that traces scaffold strands through the target structures. Moreover, unlike conventional origami designs built from close-packed helices, our structures have a more open conformation with one helix per edge and are therefore stable under the ionic conditions usually used in biological assays.

The starting point of the method we present here is a 3D mesh representing the geometry one wishes to realize at the nanoscale.

Focusing only on polyhedral meshes, that is, meshes which enclose a volume inflatable to a ball, and in contrast to several previous approaches^{14,17,19} (see Extended Data Fig. 1), we aim to replace the edges of the mesh by single DNA double helices such that the scaffold strand traverses each of these edges once. This problem is closely related to the ‘Chinese postman tour’ problem²² in graph theory, and finding solutions by hand would be impossible in practice for most meshes. The main three principles underpinning our design paradigm are: first, that the technique should allow meshes to be triangulated to optimize structural rigidity; second, that each edge should be represented by one double helix to enable construction of large structures using as little DNA as possible (though some meshes require two helices to render certain edges, as discussed below); and, third, that vertices should be non-crossing (that is, the scaffold should not cross itself at the vertices, which ensures non-knotted paths with fewer topological and kinetic traps during folding, and vertex junctions should be planar, which avoids mesh protrusions caused by the stacking of crossing helices at each vertex).

The overall design scheme is split into four discrete steps, as follows. (1) Drawing of a 3D polygon mesh using 3D software; see Fig. 1a. (2) Generating an appropriate routing of the long scaffold strand through

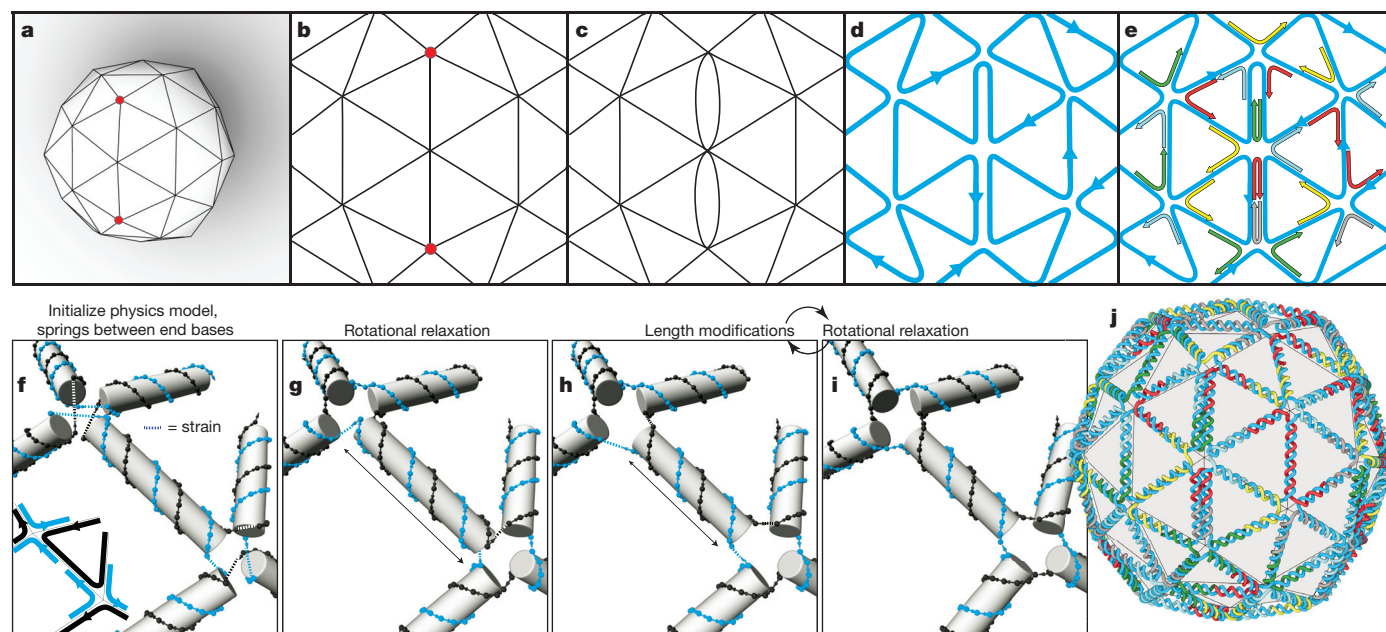


Figure 1 | Design paradigm and automated workflow for scaffold-routing sequence design of origami 3D meshes. **a**, A 3D mesh is drawn using 3D software. **b**, Using the minimum weight perfect matching algorithm, odd-degree vertices are paired. **c**, Double edges are introduced. **d**, The developed A-trails algorithm routes the scaffold according to the constraints. **e**, The staple-strand (multi-coloured) routing follows implicitly from the scaffold (blue)

routing. **f–i**, Before computation of the sequences, a physics model is used to relax and evenly distribute strain in the design. Each double helix is treated as a stiff rod with springs connecting the bases at the ends of the scaffold strand and staple strand. Iterations of rotational relaxation (**g** and **i**) and length modification of helices (**h**) leads to the final design (**j**), where sequences are calculated after importing to vHelix.

¹Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-171 77 Stockholm, Sweden. ²Department of Neuroscience, Karolinska Institutet, SE-171 77 Stockholm, Sweden.

³Department of Computer Science, Aalto University, FI-00076 Aalto, Finland. ⁴Department of Cell and Molecular Biology, Karolinska Institutet, SE-171 77 Stockholm, Sweden.

all the edges of the mesh; see Fig. 1b–e. (3) Determining the least-strained DNA helix arrangement that will realize the desired 3D mesh; see Fig. 1f–i. (4) Optional fine tuning of the design and generation of the staple strands; see Fig. 1j.

Having selected a target 3D polygon mesh (design step (1)), the first condition for a triangulated mesh to be routable (design step (2)), with the scaffold strand traversing every edge once, is that the mesh graph must admit an Eulerian circuit, that is, all its vertices have an even degree. To make meshes Eulerian, we use a general re-conditioning algorithm that adds ‘helper edges’ by introducing extra helices along certain edges; see Fig. 1b–d. A re-conditioning with the minimum number of additional helper edges, with at most one additional helper helix per edge, amounts to finding a ‘minimum weight perfect matching’ of odd-degree vertices (compare with Supplementary Fig. 1.1 in Supplementary Note 1). However, Eulerian circuits are not sufficient to ensure scaffold routing, because such circuits may generally have multiple crossings at many vertices, and even ‘elementary non-crossing’ circuits cannot always be connected by complementary staple strands (see Supplementary Fig. 1.3 in Supplementary Note 1).

These considerations lead us to adopt a routing based on A-trails²³, a specific type of Eulerian circuits, where consecutive edges of the circuit are always neighbours in the cyclic ordering around the vertices

(exemplified in Fig. 1d). Although there are efficient algorithms for finding Eulerian circuits and for minimum weight perfect matchings²², it is strongly believed that there is no efficient algorithm for finding A-trails in general graphs, or even in polyhedral graphs²³; that is, the problem is known to be NP-complete. Nevertheless, by the systematic search we developed, employing pruning and a heuristic for branching, our algorithm managed to find a routing for all the designed meshes within seconds. (For more detailed discussion of the graph theory guiding our scaffold routing procedure and the associated algorithm, see Methods and Supplementary Note 1.)

The routing of the staple strands (the helper DNA oligonucleotides that drive DNA-origami folding) follows implicitly from completing the edge connections at the vertices (Fig. 1e). For design step (3), a physical model of rigid cylinders joined by stressed springs at the vertex junctions is implemented *in silico* and allowed to relax in a simulation to give fewer overlaps and smaller gaps at the vertices. In the case of helices where the routing complicates the rotational relaxation (that is, where connections on the opposite end of the helix try to rotate the helix in the opposite direction; see Fig. 1g), we add an iterative length-modification step in the relaxation algorithm (Fig. 1h) to adjust the lengths of individual edges. (See Methods and Supplementary Note 2 for a full description of the

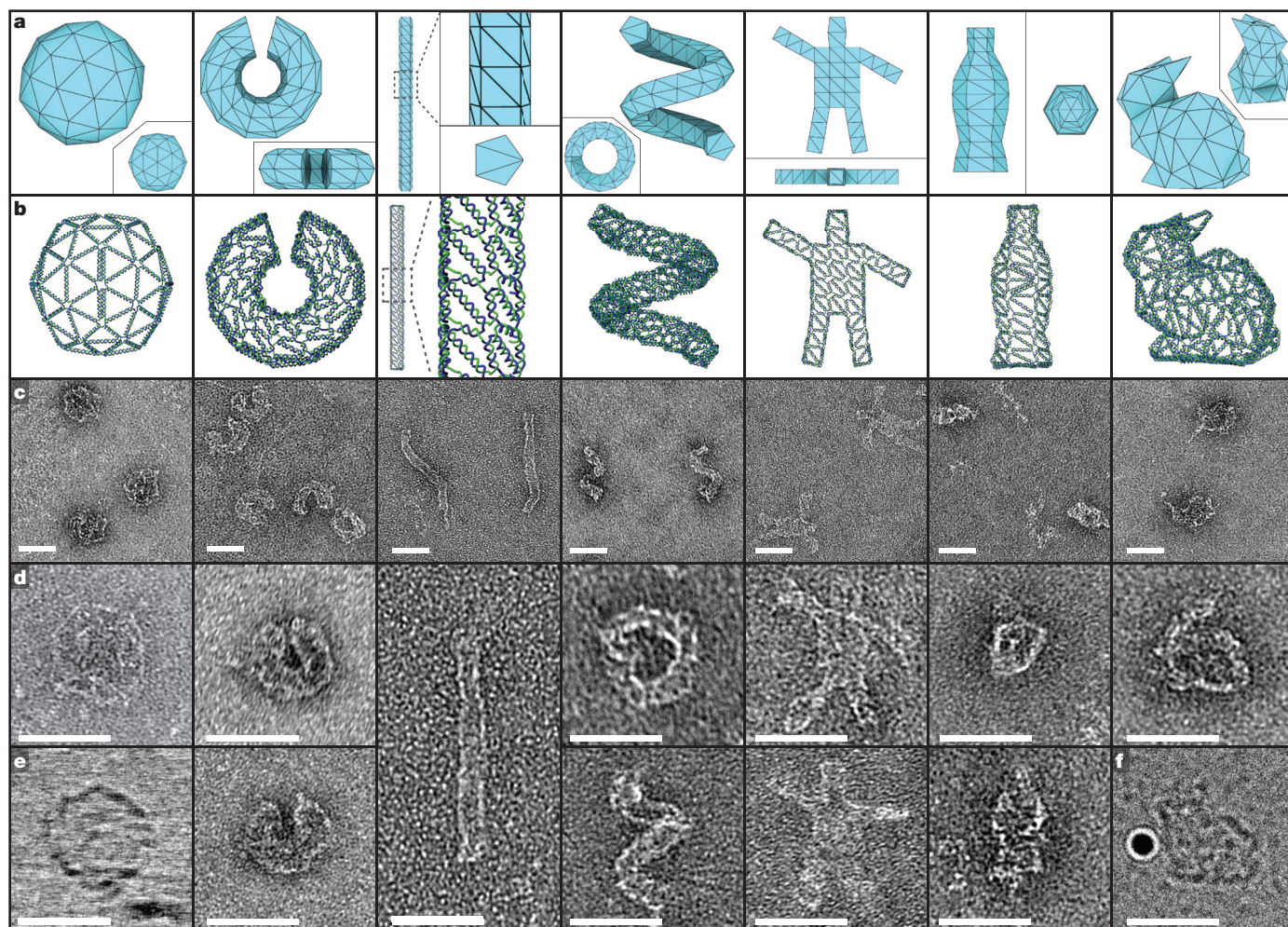


Figure 2 | 3D meshes rendered in DNA. **a**, Different views of the 3D meshes provided as starting points for the automated design process. In columns from left to right: a ball generated by subdivision of an icosahedron, a nicked torus, a rod and a helix with pentagonal cross-sections, a thin, semi-two-dimensional, waving stickman, a bottle and a version of the Stanford bunny. **b**, The front face of the complete DNA designs in each case with single DNA strands rendered as tubes: the staple strands in blue and the scaffold strands in

green. **c–e**, Negative-stain dry-state transmission electron microscopy (except for the ball and bunny in **e** and **f**, respectively) micrographs of each of the structures. **c**, 250 nm × 250 nm views. **d** and **e** show 100 nm × 100 nm close-ups (excluding the pentagonal rod, which is 200 nm × 100 nm). **e**, **f**, The ball and bunny are imaged using cryo-electron microscopy (the gold particle used for alignment is visible in **f**). Scale bars are 50 nm.

physical relaxation simulation and length-correction process.) The relaxed model is then imported into Autodesk Maya running vHelix (<http://www.vhelix.net>), a custom-made dedicated plugin for the design and visualization of DNA nanostructures, as shown in Extended Data Fig. 2.

For fine-tuning of the design (optional step (4)), the smaller gaps within the imported relaxed model can be filled with unpaired nucleotides, which will provide flexibility and correct strand misalignments during assembly. If desired, further manual post-processing of the design, such as modifying staple-strand breakpoints, can be done in vHelix. Finally, we introduce the desired scaffold-strand sequence, and then vHelix automatically generates the staple-strand sequences, thus completing the design process.

Overall, the set of tools provided allows a target 3D geometry to be rendered with DNA automatically, with fine-grained control over the design in a graphical user interface before sequence generation. An outline view of the complete pipeline is given in Extended Data Fig. 3.

We designed six polyhedral models in Autodesk Maya: a ball, a nicked torus, a helix, a rod, a humanoid stickman and a soda bottle. From a downloaded and imported model (<http://visual.k.u-tokyo.ac.jp/research/unfolding/index-e.html>), we also produced a reduced polygon version of the Stanford bunny. We scaled the structures to scaffold sizes of between six and eight thousand nucleotides. The scaling (physical dimension) of each model can be set arbitrarily before the relaxation simulation and will determine the double-helix characteristics at each edge. Implicitly, the scaling also affects the number of edges, or 'resolution', of a polyhedral model that can be rendered from a given strand of DNA of a certain length. That is, the number of edges, combined with the overall size of the object, determines how long the DNA scaffold must be.

The routing of the staple strands is fully determined by the scaffold routing. However, the placement of staple-strand breakpoints can be freely modified. In the case of the symmetrical ball structure, we designed the breakpoints using a simple scheme in which each staple strand attaches to two adjacent half-edges of the routed scaffold. In the other structures, which have a larger spread of edge lengths, we implemented an automatic scheme for staple-strand breakpoint design in which staple strands were designed to hybridize (pair) with more than two edges. This avoided breakpoints on the shortest edges, which allowed them to be scaled to a smaller size.

We found in the vHelix models that in some of our target structures, which are complex and strongly curved 3D objects, some of the strands in certain vertices appeared to leave gaps in the junctions and could lead to strain in the final assembly. We therefore implemented a feature in vHelix that relaxes such strands through the addition of extra unpaired bases on either the scaffold strands or on the staple strands in the vertices. When the unpaired bases were placed on the staple strands, we designed them to be adenines. This was implemented for all structures except the ball.

All structures were folded at an excess of 10× staple strands to scaffold strands and evaluated in agarose gel electrophoresis (Supplementary Figs 1–7). Subsequent imaging of the DNA structures in negative-stained transmission electron microscopy (Fig. 2c and d; Supplementary Figs 8–14) revealed objects in good accordance with prediction, although the hollow structures sometimes appear collapsed in the dried-out state of negative-stained transmission electron microscopy. Cryo-electron tomography however, allows imaging of the structures in a hydrated state (Fig. 2e, f, Fig. 3 and Supplementary Figs 15 and 16) and revealed that they did indeed fold into their desired shapes; the tomography even uncovered, in close up, features of the underlying DNA mesh (Fig. 3).

Most DNA origami structures are built from tightly packed helices stabilized by multivalent cations or high concentrations of monovalent cations²⁴, preventing them from folding and remaining stable in physiological buffer systems. Our new polyhedral structures do not share the close packing of helices, and we found that the ball, helix,

nicked torus and stickman folded and remained stable in two buffers commonly used in biomedical research—phosphate buffered saline (PBS) and Dulbecco's Modified Eagle Medium (DMEM)—and when using a classic magnesium-rich buffer (Fig. 4 and Supplementary Figs 17–25). We see some evidence of aggregation when folding in DMEM, which might be alleviated by folding in PBS followed by buffer exchange into DMEM. Compared to standard origami²⁵, the structures appear more stable in cell culture buffers supplemented with non-inactivated serum (fetal bovine serum) but exhibit similar sensitivity to high concentrations of nuclease (Supplementary Figs 25 and 26; see also Methods).

The ease of the design process is an important parameter for determining whether a new nanotechnology method will find wide use. Because our method is highly amenable to full automation, it opens up the possibility of 'one-click' 3D printing at the nanoscale: the user would draw a polygonal shape in 3D software and would then be directly provided with the DNA sequences to order. The ball, bottle and bunny designs were all generated in such a completely automated fashion, giving DNA structures directly from digital 3D meshes without manual intervention. The paradigm renders structures that fold well with a yield of 5%–92% estimated from agarose gel electrophoresis (Supplementary Figs 1–7, Supplementary Table 1) and where almost all particles examined (from the leading agarose gel band) in cryo-electron tomography appear well formed (Figs 2 and 3; Supplementary Figs 15 and 16).

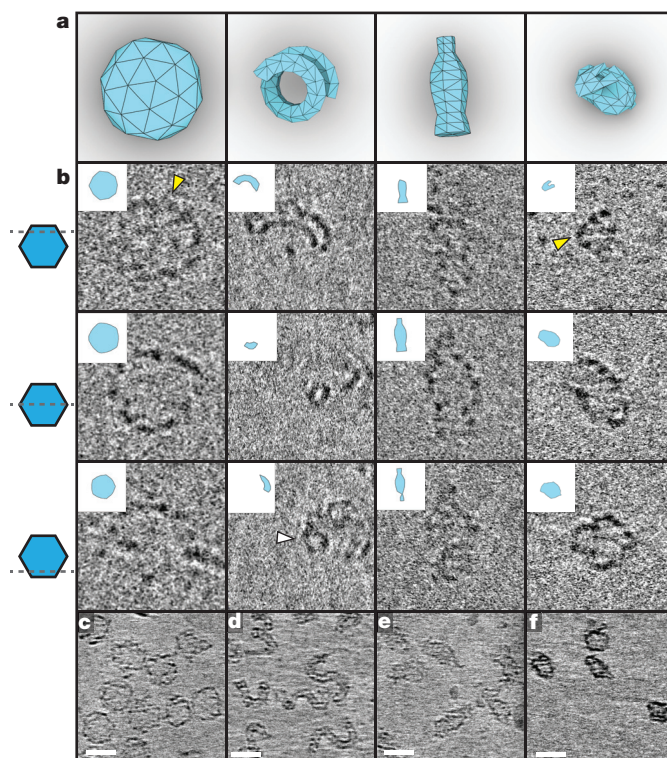


Figure 3 | Cryo-electron microscopy reveals the hollow characteristics and details of polyhedral meshes. From left to right, the ball, helix, bottle and bunny are shown. **a**, 3D renderings of the structures shown in the cryo-electron tomography images in **b** rotated to correspond to the particles observed in the data. **b**, Three progressive slices of the structures reconstructed from cryo-electron tomography imaging. Images are 100 nm × 100 nm wide. Insets show the expected outlines from the corresponding sections of the digital models. Mesh triangulation can be observed (yellow arrows), as well as the pentagonal cross-section of the helix tube (white arrow). **c–f**, Overview images from cryo-electron microscopy of each of the structures (contrast projection reconstructions, obtained by averaging multiple tomography slices). Scale bars are 50 nm.

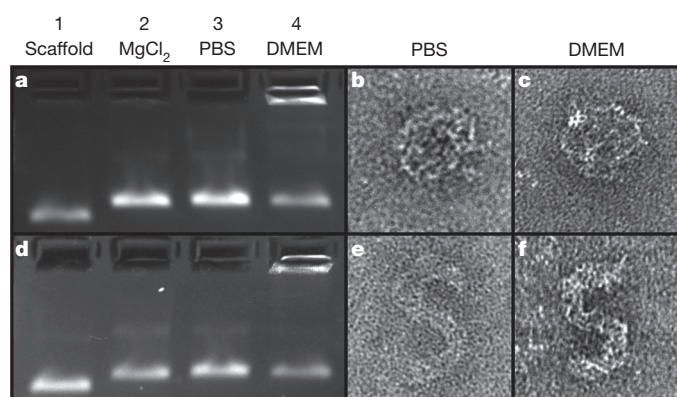


Figure 4 | Mesh origami folds in and is fully stable in physiological buffers. **a, d,** Agarose gel electrophoresis of the ball (**a**) and the helix (**d**) folded in different buffers. Scaffold DNA is shown in lane 1. Structures are folded in cell media as follows: in 10 mM MgCl₂ and 5 mM Tris (lane 2), in PBS (lane 3) and in DMEM (lane 4). **b, c, e, f,** Transmission electron micrographs of the ball (**b** and **c**), and the helix (**e** and **f**) folded in PBS (**b** and **e**) or DMEM (**c** and **f**). Images are 100 nm × 100 nm wide.

This work is the first to base DNA origami architecture on A-trails routing theory. But the rational design of small-protein nanostructures using other types of Eulerian paths has also been recently reported²⁶, further highlighting the value of a deeper mathematical understanding of path routing in the self-assembly of linear molecules. In this case, by exactly formulating the non-crossing scaffold routing problem as a search for a specific type of Eulerian circuits in polyhedral graphs and then connecting this search problem to a long-standing conjecture in graph theory concerning the existence of A-trails in planar Eulerian triangulations, we arrived at an effective branch-and-bound search algorithm that makes it feasible to find the requisite scaffold routings quickly, even in 3D designs with a large polygon count, despite the problem being NP-complete.

We hypothesize that the open folding architecture we present could be particularly well suited for folding using very long staple strands²⁷ for increased thermal stability in the future. A similar long-staple-strand strategy is believed to be difficult to implement using normal origami routing owing to the intrinsically high degree of topological complexity.

3D DNA origami has traditionally been implemented using close-packed helices that can yield solid brick-like shapes^{14,15} that are both impressive and visually appealing when imaged using dry-state negative-stain transmission electron microscopy. But emerging work^{28–30} that utilizes DNA origami in biological research, where qualities such as stability in low-salt conditions and structural flexibility are important, has favoured one-layer, hollow structures. The new design paradigm we report here, using double helices alone as structural elements instead of close-packed bundles of helices, alleviates the need for non-physiological concentrations of salts completely, and is expected to enable more experiments in cell biology and potentially also *in vivo*, with a closer match between conditions in the model system and the true biological context.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 February; accepted 14 May 2015.

- Seeman, N. C. Nucleic-acid junctions and lattices. *J. Theor. Biol.* **99**, 237–247 (1982).
- Rothmund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
- Chen, J. H. & Seeman, N. C. Synthesis from DNA of a molecule with the connectivity of a cube. *Nature* **350**, 631–633 (1991).

- Rothmund, P. W. K., Papadakis, N. & Winfree, E. Algorithmic self-assembly of DNA Sierpinski triangles. *PLoS Biol.* **2**, e424 (2004).
- He, Y., Tian, Y., Ribbe, A. E. & Mao, C. Highly connected two-dimensional crystals of DNA six-point-stars. *J. Am. Chem. Soc.* **128**, 15978–15979 (2006).
- Shih, W. M., Quispe, J. D. & Joyce, G. F. A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron. *Nature* **427**, 618–621 (2004).
- Goodman, R. P. *et al.* Rapid chiral assembly of rigid DNA building blocks for molecular nanofabrication. *Science* **310**, 1661–1665 (2005).
- Bhatia, D. *et al.* Icosahedral DNA nanocapsules by modular assembly. *Angew. Chem. Int. Edn Engl.* **48**, 4134–4137 (2009).
- Geary, C., Rothmund, P. W. K. & Andersen, E. S. A single-stranded architecture for cotranscriptional folding of RNA nanostructures. *Science* **345**, 799–804 (2014).
- List, J., Weber, M. & Simmel, F. C. Hydrophobic actuation of a DNA origami bilayer structure. *Angew. Chem. Int. Edn Engl.* **53**, 4236–4239 (2014).
- Andersen, E. S. *et al.* Self-assembly of a nanoscale DNA box with a controllable lid. *Nature* **459**, 73–76 (2009).
- Marchi, A. N., Saaem, I., Vogen, B. N., Brown, S. & LaBean, T. H. Toward larger DNA origami. *Nano Lett.* **14**, 5740–5747 (2014).
- Ke, Y. *et al.* Scaffolded DNA origami of a DNA tetrahedron molecular container. *Nano Lett.* **9**, 2445–2447 (2009).
- Douglas, S. M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009).
- Dietz, H., Douglas, S. & Shih, W. M. Folding DNA into twisted and curved nanoscale shapes. *Science* **325**, 725 (2009).
- Han, D. *et al.* DNA gridiron nanostructures based on four-arm junctions. *Science* **339**, 1412–1415 (2013).
- Iinuma, R. *et al.* Polyhedra self-assembled from DNA tripods and characterized with 3D DNA-PAINT. *Science* **344**, 65–69 (2014).
- Liedl, T., Högberg, B., Tytell, J., Ingber, D. E. & Shih, W. M. Self-assembly of three-dimensional prestressed tensegrity structures from DNA. *Nature Nanotechnol.* **5**, 520–524 (2010).
- Rothmund, P. W. K. in *Nanotechnology: Science and Computation* (eds Chen, J., Jonoska, N. & Rozenberg, G.) 3–21 (Springer, 2006).
- Ellis-Monaghan, J. A., McDowell, A., Moffatt, I. & Pangborn, G. DNA origami and the complexity of Eulerian circuits with turning costs. *Natural Comput.* <http://dx.doi.org/10.1007/s11047-014-9457-2> (2014).
- Douglas, S. M. *et al.* Rapid prototyping of 3D DNA-origami shapes with caDNANO. *Nucleic Acids Res.* **37**, 5001–5006 (2009).
- Edmonds, J. & Johnson, E. L. Matching, Euler tours and the Chinese postman. *Math. Program.* **5**, 88–124 (1973).
- Davling Andersen, L. & Fleischner, H. The NP-completeness of finding A-trails in Eulerian graphs and of finding spanning trees in hypergraphs. *Discrete Appl. Math.* **59**, 203–214 (1995).
- Martin, T. G. & Dietz, H. Magnesium-free self-assembly of multi-layer DNA objects. *Nature Commun.* **3**, 1103 (2012).
- Hahn, J., Wickham, S. F. J., Shih, W. M. & Perrault, S. D. Addressing the instability of DNA nanostructures in tissue culture. *ACS Nano* **8**, 8765–8775 (2014).
- Gradišar, H. *et al.* Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nature Chem. Biol.* **9**, 362–366 (2013).
- Ducani, C., Kaul, C., Moche, M., Shih, W. M. & Högberg, B. Enzymatic production of ‘monoclonal stoichiometric’ single-stranded DNA oligonucleotides. *Nature Methods* **10**, 647–652 (2013).
- Shaw, A. *et al.* Spatial control of membrane receptor function using ligand nanocalipers. *Nature Methods* **11**, 841–846 (2014).
- Douglas, S. M., Bachelet, I. & Church, G. M. A logic-gated nanorobot for targeted transport of molecular payloads. *Science* **335**, 831–834 (2012).
- Amir, Y. *et al.* Universal computing by DNA origami robots in a living animal. *Nature Nanotechnol.* **9**, 353–357 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was funded through grants from the Swedish Research Council (grants 2010-5060 and 2013-5883 to B.H.), the Swedish Foundation for Strategic Research (grant FFL12-0219 to B.H.) and the Knut and Alice Wallenberg foundation (Academy Fellow grant KAW2014.0241 to B.H.). E.B. was also funded by a Wallenberg Scholars grant to O. Inganäs. The work of A.M. was supported by the Helsinki Doctoral Education Network on Information and Communications Technology. We acknowledge the computational resources provided by the Aalto Science-IT project and the use of the Facility for EM Tomography at the Karolinska Institutet. A.M., E.C. and P.O. thank G. García Pérez for help in implementing early variants of the scaffold routing software. We thank P. Kumar Aredy (Karolinska Institutet) and A.-L. Bank Kodal and K. Gothelf (Aarhus University) for help with early experimental testing.

Author Contributions B.H. conceived and designed the study. P.O., E.C. and A.M. performed the theoretical work. E.B. performed the experimental work. A.M. performed most of the algorithm implementation. J.G. collaborated on the algorithm implementation and implemented the relaxation method and the vHelix plug-in. S.M. performed the cryo-electron tomography and tomography with help from E.B. All authors contributed to the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.H. (bjorn.hogberg@ki.se).

METHODS

Code availability. Links to source code for all software, instructional videos and tutorials that pertain to the polygonal design process reported can be found at <http://www.vhelix.net>.

Mesh design. The 3D objects were designed in Autodesk Maya 2014 as polygon objects. The feature “triangulate” was used to triangulate the meshes. Note that fully triangulated convex polyhedral meshes are structurally rigid³¹. When designing meshes, the number of edges was kept to an appropriate amount for the intended scaffold in the downstream design and the spread in edge lengths was kept small.

The routing algorithm works on meshes in ASCII PLY format. Autodesk Maya does not currently support export in this format, so meshes were exported in the STL format and converted to PLY using the tool meshconv (<http://www.cs.princeton.edu/~min/meshconv/>).

Routing and relaxation. The algorithms for reconditioning and routing of the mesh as well as physically relaxing it for import to vHelix are all performed by running a single BAT file on the mesh file using the command prompt:

```
bcor.bat model_ply_file scale
```

The decimal value *scale* is used in the physical relaxation where the edges of the mesh are converted into DNA helices of a certain length. The batch script generates a scaffold routing of the mesh given in PLY file format by executing a sequence of modules. Note that the precise formulation of the scaffold routing problem and a detailed discussion of the graph-theoretic concepts applied in the modules are available in Supplementary Note 1.

The first module converts the PLY file format into the DIMACS format, which is a widely used representation for graphs. Note that the 3D positioning of the vertices is lost in this conversion. Nevertheless, the relative—that is, cyclic (for example, counter-clockwise)—order of edges around vertices, which is essential information when finding an A-trail routing, can still be obtained by a planar embedding algorithm in the third step.

The second module applies the reconditioning of the graph if the graph is not Eulerian. The scaffold is able to traverse each edge exactly once if and only if the graph is Eulerian. Hence, once-per-edge scaffold routing is only possible if there are no odd-degree vertices³². Thus, the module identifies the odd-degree vertices and (if such vertices exist) applies Edmonds’ blossom algorithm²² for minimum weight perfect matching. The result of the matching yields a transformation of the initial graph into an Eulerian one by adding a minimum amount of double edges between pairs of odd-degree vertices. If on the other hand there are no odd-degree vertices, the script continues with the original graph.

The third module applies the Boyer–Myrvold algorithm³³ and generates a planar embedding of the graph. Since the input graph is from a polyhedral mesh, it is 3-connected³⁴, that is, the removal of any two vertices does not leave the graph disconnected. Hence, by Whitney’s unique embedding theorem³⁵, the generated embedding retains the cyclic order of edges around vertices in the 3D mesh.

The first three modules essentially prepare the mesh for routing. As detailed in Supplementary Note 1, the desired form of routing is based on A-trails, where for polyhedral graphs consecutive edges in the circuit always lie on the same face boundary³⁶. However, the search for A-trails is expensive, with the problem known to be NP-complete both for general planar graphs³⁷ and polyhedral graphs²³.

Nevertheless, the last module performs an A-trail search on the embedding based on a systematic branch-and-bound search. The algorithm constructs the search tree based on binary choices for vertices of degree at least six, resulting in an Eulerian circuit in a derived graph with maximum degree four. All the crossings of the circuit that are at degree-four vertices in the derived graph can then be removed in polynomial time³⁸. The structuring of the search tree, coupled with a heuristic for branching order, enabled the generation of routings for large meshes such as those designed in this work. For instance, the routing for the Stanford bunny, the most complex among the ones designed, was obtained in 87.87 ms (average of 11 runs) on a midrange workstation (Intel Core i5-2500 CPU at 3.30 GHz, 8 GB RAM, Windows 7 64-bit OS). The software code of the modules for the routing is freely available online at <https://github.com/moham1/bcor>.

In the last step before the import into vHelix, the polyhedral mesh is converted into a DNA design of a discrete size. Here, adequate strand lengths for all edges as well as the position of the helices need to be determined. This is done by iteratively minimizing the overall structural tension, as described below and in more detail in Supplementary Note 2.

To find the optimal translation and orientation of helices along the edges, the placement of these is simulated in a spring–rigid-body setup. By approximating the initially placed DNA helices as rigid-body cylinders and the connectivity between endpoint nucleotides of different helices as spring-joints, the total

accumulated separation energy of these can easily be minimized by any rigid-body physics simulation engine. We used the Nvidia PhysX engine (<https://developer.nvidia.com/physx-sdk>).

In the first iteration, the routed structure is loaded and the length of the rigid edges is discretized as a multiple of base-pair lengths (0.33 nm) given by the mesh size and a user-selected scaling factor. Then the relaxation simulation is run and strain on the connecting springs is calculated.

The relaxation optimization is implemented as an iterative process where the physics simulation, described above, alternates with a length-modification step. In this step, one edge is shortened or lengthened by one base. After this, the simulation is started anew and run until a new minimum is found. If this new minimum is a better fit than the previous one, it replaces the current structure in the search for further modifications. If not, the modification will be discarded and the algorithm will modify another helix to attempt to find a lower accumulated spring energy. After the algorithm has been unable to further successfully optimize the structure, the rotation, translation and length of the helices is extracted from the simulation and joined with the routing to produce an output file in the format .rply. This file can be imported to vHelix for further manipulation and design of the origami structure.

On a midrange workstation, the physical relaxation may take hours to fully complete for some structures. However, it may be interrupted by the user and will then output the latest state for import to vHelix. For the Stanford bunny the relaxation took an average of 20 min 35 s (based on 11 runs) on a midrange workstation (Intel Core i5-2500 CPU at 3.30 GHz, 8 GB RAM, Windows 7 64-bit OS).

vHelix. vHelix is our custom-made plug-in for Autodesk Maya. The plugin allows a user to manipulate a model of DNA in 3D and connect strands together freely. The DNA model is programmed to closely emulate known DNA geometry and Holliday junctions created in vHelix closely recreates crystallography data of DNA³⁹. vHelix is used to inspect the final design, and allows the user to make manual edits directly in the 3D model if desired.

When Autodesk Maya is running with the vHelix plug-in, .rply files can be directly imported using the import menu in Maya. In the import, the staple-strand breakpoints are positioned. If the import menu is used, vHelix will use its default method and position breakpoints at the middle of each edge creating staple strands that each bind to two half-edges. This is used for the ball structure. To achieve a more sophisticated positioning of the breakpoints the .rply file can be imported to vHelix using a MEL command in the Maya script editor:

```
file -options “nicking_min.length = x;nicking_max.length = y” -import -type
“Text based vhelix” -ra true -mergeNamespacesOnClash false -namespace
“file name” -pr “file path”
```

where x , y , “file name” and “file path” are variables that the user should change to what is applicable for that particular import. The integer x controls the minimum-edged length at which breakpoints should be positioned. In meshes with a large spread in edge sizes it is appealing to position the breakpoints on the longer edges. The integer y controls the maximum allowed staple-strand length; this limit is normally motivated by the length of oligonucleotides that can be synthesized inexpensively. Often, it is not possible to satisfy both parameters in one structure and the result may be a compromise violating one or both parameters. By using the vHelix feature “export strands” the lengths of the staple strands created can be evaluated. If no sequence is assigned to the structure, the exported file gives a list of question-marked sequences that correspond to the undetermined base sequences of the staple strands together with a name describing the helices connected by the staple strand. The helix names can be found in the outliner of Maya for easy inspection of staple strands. This was used for all structures except the ball. In our experience, the simpler staple-strand design is preferable whenever possible and multi-edge stapling should primarily be used to avoid nicks on very short edges.

If the staple strands generated are not satisfactory, the auto-breakpoint design can be rerun with other parameters or the staple strands can be manually remodelled. Staple strands that are too long can be shortened by the manual introduction of breakpoints by selecting a base at the position of the desired breakpoint and using the feature “disconnect bases”. Breakpoints can also be removed by selecting the two adjacent bases and using the feature “connect bases”. Using these two features, the breakpoints of the staple strands can be manually remodelled. If no automatic breakpoints are desired the .rply file can be imported using large values for x and y . This will generate staple strands with the maximal possible length as determined by the scaffold routing with the possibility of manually introducing staple-strand breakpoints. For the rod, nicked torus, helix and stickman, some manual modifications were made to the breakpoints. For the bottle and Stanford bunny, no manual modifications were made.

In vHelix the scaffold and staple strands may appear visually as though they are nicked at junctions although they are actually connected. This may induce stress in the folded structures because the junctions may be more tightened than in the vHelix representation. This can be countered by the feature “auto fill strand gaps” in vHelix. The feature will iterate over the DNA strands to search for gaps between bases. If a gap is found, it will be filled by the addition of extra unpaired bases. This can be performed on a single selected strand or on all staple and scaffold strands if no strand is selected. This was used for all structures except the ball; adenine was used to supply unpaired extra bases. We designed a version of the rod without this feature and found that it did not fold successfully, possibly owing to high tension in the junctions. Therefore we recommend the use of this feature in most designs.

To generate staple-strand sequences the sequence of the scaffold strand is assigned by selecting a base of the scaffold and using the feature “apply sequence”. This will automatically assign the complementary sequence to the staple strands. If the used scaffold is longer than needed for the structure, the excess unpaired bases will form a loop at the position selected. This may affect the structural stability and so the position of the loop should be chosen to be where the loop will not interfere with other parts of the structure. After sequences have been assigned, the feature “export strands” can be used to export the sequences of a selected strand or all strands if no selection is made. The sequences are exported in a comma-separated file that can be easily imported to a spreadsheet application. If the feature “auto fill strand gaps” has been used, the extra unpaired bases inserted to the staple strands will appear as question marks in the exported file. They can be converted to a desired nucleotide before staple-strand ordering. For an experienced user, the design process in vHelix can be completed in less than an hour.

Scaffold DNA preparation. *Escherichia coli* strain JM109 was streaked on a lysogeny broth agar plate and grown overnight at 37 °C to produce separate colonies. A single colony was cultured overnight in 25 ml lysogeny broth as a pre-culture. 3 ml of this culture was diluted in 250 ml of 2× Yeast extract Tryptone growth medium (Sigma Aldrich) medium with 5 mM MgCl₂ (VWR International) and placed in a shaker at 37 °C. During growth, the optical density at 600 nm was measured repeatedly until it reached 0.5. Then the phage for scaffold was added at a multiplicity of infection of 1, and incubation with shaking continued for an additional 4 h. The culture was transferred to a 250-ml centrifuge bottle and was centrifuged at 4,000g for 30 min to pellet the bacteria, and the supernatant containing the phage was centrifuged again at 4,000g for 20 min. 10 g of PEG 8000 (Amresco) and 7.5 g of NaCl (VWR International) were added to the supernatant, which was then incubated on ice for 30 min and centrifuged at 10,000g for 40 min to pellet the phage. Next, the supernatant was discarded, and the pellet was resuspended in 10 ml of 10 mM Tris (pH 8.5, VWR International) and transferred to a 85-ml centrifuge bottle. 10 ml of a solution with 0.2 M NaOH (VWR International) and 1% SDS (VWR International), was added, mixed gently by inversion and incubated at room temperature for 3 min. Then 7.5 ml of 3 M KOAc (VWR International), pH 5.5, was added, gently mixed by swirling and incubated on ice for 10 min to denature the phage protein coat. The mixture was centrifuged at 16,500g for 30 min to pellet the phage protein. The supernatant containing DNA was poured into fresh centrifuge bottles, and 50 ml 99.5% EtOH (Kemetyl) was added, mixed gently by inversion and incubated on ice for 30 min and then centrifuged at 16,500g for 30 min to precipitate the DNA. The supernatant was carefully discarded and the pellet was washed with 75% EtOH and air dried at room temperature for 15 min. The pellet was resuspended in 2 ml of 10 mM Tris, pH 8.5, and the concentration and quality were characterized by ultraviolet–visual spectroscopy (NanoDrop, Thermo Scientific) and a 2% agarose gel, respectively.

Staple oligonucleotide preparation. Staple oligonucleotides were purchased from Integrated DNA Technologies. They were delivered desalted in water in 96-well plates at a concentration of 100 µM each. The staple strands were pooled and diluted with water to a working concentration of 400 nM each. Lists of staple-strand sequences are found in Supplementary Tables 2–8.

Folding. In the folding reactions, the scaffold DNA was diluted to 5 nM and the staple strands diluted to 50 nM each, corresponding to a 10× excess of each staple to the scaffold. Scaffold-strand M13mp18 was used for the ball, nicked torus, stickman and bottle, p7560 was used for the helix and bunny, and p8064¹⁵ was used for the rod. For standard folding the mix was brought to 5 mM Tris 1 mM ethylenediaminetetraacetic acid (EDTA, VWR International) and between 4 mM and 10 mM MgCl₂. For folding in PBS the sample was mixed with 10× PBS (Sigma Aldrich) to 1× PBS in the final mix. For folding in DMEM the sample was mixed with 10× DMEM (Sigma Aldrich) supplemented with sodium bicarbonate (Sigma

Aldrich), as instructed by the manufacturer. The mixed sample was put on a thermal ramp starting with a rapid heat denaturation at 80 °C for 5 min followed by cooling from 80 °C to 60 °C over 20 min, then slow cooling from 60 °C to 24 °C over 14 h.

Agarose gel electrophoresis. Agarose gels were cast using 2% agarose (VWR International) in 0.5× Tris/borate/EDTA (TBE buffer) supplemented with 10 mM MgCl₂ and 0.5 mg ml^{−1} ethidium bromide (Sigma Aldrich). Gels were run in 0.5× TBE buffer supplemented with 10 mM MgCl₂ on ice at 70 V for 4 h on ice. After running, gels were imaged in a GE LAS 4000 imager.

Gel extraction of structures for electron microscopy. Samples were run in 0.8% agarose gels with 0.5× TBE buffer supplemented with 10 mM MgCl₂ and 0.5 mg ml^{−1} ethidium bromide at 90 V until adequate separation was achieved. The band containing well-folded structures was cut out and smashed using a micro-pestle. The smashed band was transferred to a freeze and squeeze gel extraction column (Bio-Rad) and centrifuged at 13,000g for 3 min.

Negative-stain transmission electron microscopy. A 400-µl aliquot of 2% w/v uranyl formate (Electron Microscopy Sciences) was mixed with 8 µl of 1 M NaOH and centrifuged at 16,500g for 5 min. 3 µl of sample was put on a glow-discharged, carbon-coated, formvar resin grid (Electron Microscopy Sciences) for 20 s before blotting on a filter paper. The sample was spotted in water and blotted again before spotting in the uranyl formate solution for 20 s. After blotting again the sample was air-dried and imaged in a FEI Morgagni 268 at 28,000× magnification.

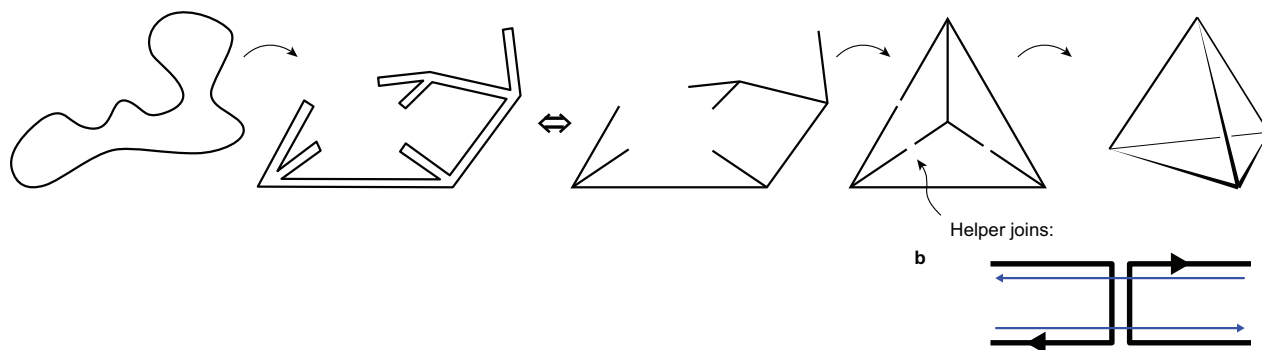
Cryo-electron tomography. Vitrobot Mk2 (FEI) was used to prepare cryo-specimens for electron microscopy/tomography. 10-nm protein-A-coated gold nanoparticles were applied as fiducial markers for image alignment to Quantifoil R2/2 grids with an additional layer of continuous carbon film after glow-discharge treatment for 20 s. The grids were additionally glow-discharged during the 2 min immediately before application of 3 µl of the sample solution. The grids were incubated at relative humidity of 90% to 100% for one to five minutes and frozen in liquid ethane after blotting (blotting time 2–3 s, drain time 1 s). The grids were transferred into a GATAN 626 cryo-holder and examined in an FEI CM200 FEG microscope under low-dose conditions.

EMMENU software (<http://www.tvips.com>) was used for automated collection of the tilt series in the range of −64° to +64° with a 4° increment. The images were recorded with a TVIPS TemCam F214 charge-coupled device camera at either 6 µm or 9 µm underfocus and a total magnification of 57,000× (pixel size equal to 4.2 Å). The dose used for image acquisition was approximately 2 electrons per Å² per image. 3D reconstruction was performed using the IMOD package⁴⁰. Two cycles of SIRT refinement (simultaneous iterative reconstructive technique) were applied to increase reconstruction quality.

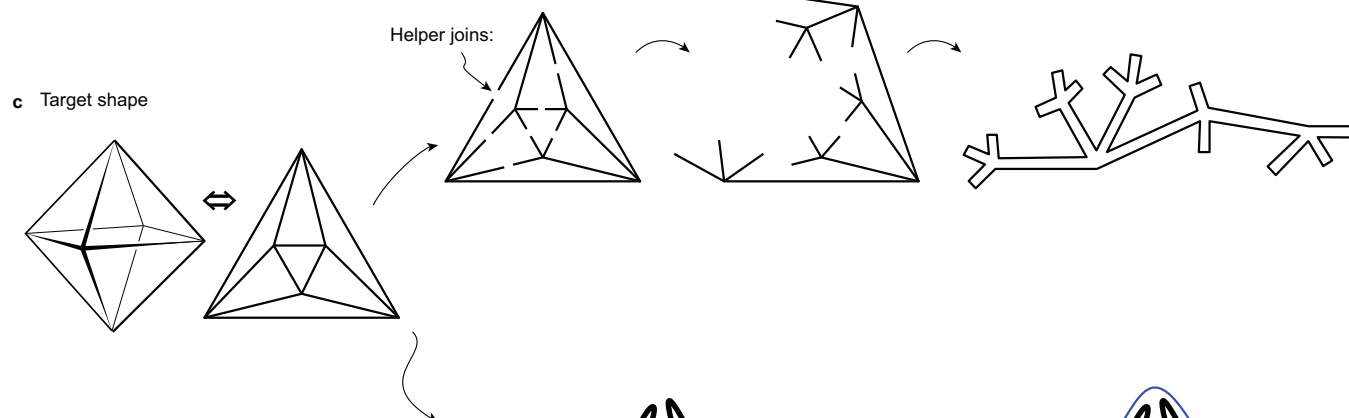
DNase stability assay. The ball and helix structures were folded in 10 mM MgCl₂, 5 mM TRIS, 1 mM EDTA as described. The structures were washed into 1× PBS supplemented with 2.5 mM MgCl₂ and 0.1 mM CaCl₂ (Sigma Aldrich) three times using a 100 kDa MWCO spin filter (Millipore). DNase I (New England Biolabs) was diluted in the same buffer and added at concentrations of 0.9–57.6 U ml^{−1} (3.6 U ml^{−1} is the average concentration in human blood⁴¹). The samples were incubated for one or twelve hours at 37 °C and then immediately loaded in a 2% agarose gel supplemented with 10 mM MgCl₂ and run for 3 h at 90 V. Gel data (Supplementary Fig. 26) indicates that the structures are stable up to 28.8 U ml^{−1} for one hour and only minor degradation is observed in samples under physiological conditions up to 12 h.

31. Cromwell, P. R. *Polyhedra* 476 (Cambridge Univ. Press, 1999).
32. Jungnickel, D. *Graphs, Networks and Algorithms* Vol. 8, 695 (Springer, 2012).
33. Boyer, J. M. & Myrvold, W. J. On the cutting edge: simplified $O(n)$ planarity by edge addition. *J. Graph Algorithms Appl.* **8**, 241–273 (2004).
34. Ziegler, G. M. *Lectures on Polytopes* 370 (Springer, 1995).
35. Fleischner, J. & Fleischner, H. *Eulerian Graphs and Related Topics* Part 1, Vol. 1, 407 (Elsevier, 1990).
36. Andersen, L. D., Fleischner, H. & Regner, S. Algorithms and outerplanar conditions for A-trails in plane Eulerian graphs. *Discrete Appl. Math.* **85**, 99–112 (1998).
37. Bent, S. W. & Manber, U. On non-intersecting Eulerian circuits. *Discrete Appl. Math.* **18**, 87–94 (1987).
38. Tsai, M.-T. & West, D. B. A new proof of 3-colorability of Eulerian triangulations. *Ars Math. Contemp.* **4**, 1 (2011).
39. Ortiz-Lombardia, M. et al. Crystal structure of a DNA Holliday junction. *Nature Struct. Biol.* **6**, 913–917 (1999).
40. Kremer, J. R., Mastrorade, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
41. Cherepanova, A. et al. Immunochemical assay for deoxyribonuclease activity in body fluids. *J. Immunol. Methods* **325**, 96–103 (2007).

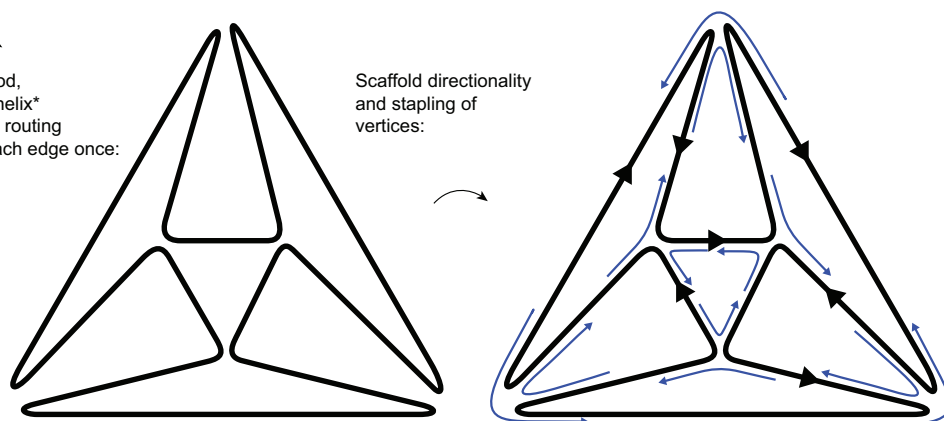
a Previously, circular scaffold folding into a tree-like shape, followed by helper joins:



d Previous methods:

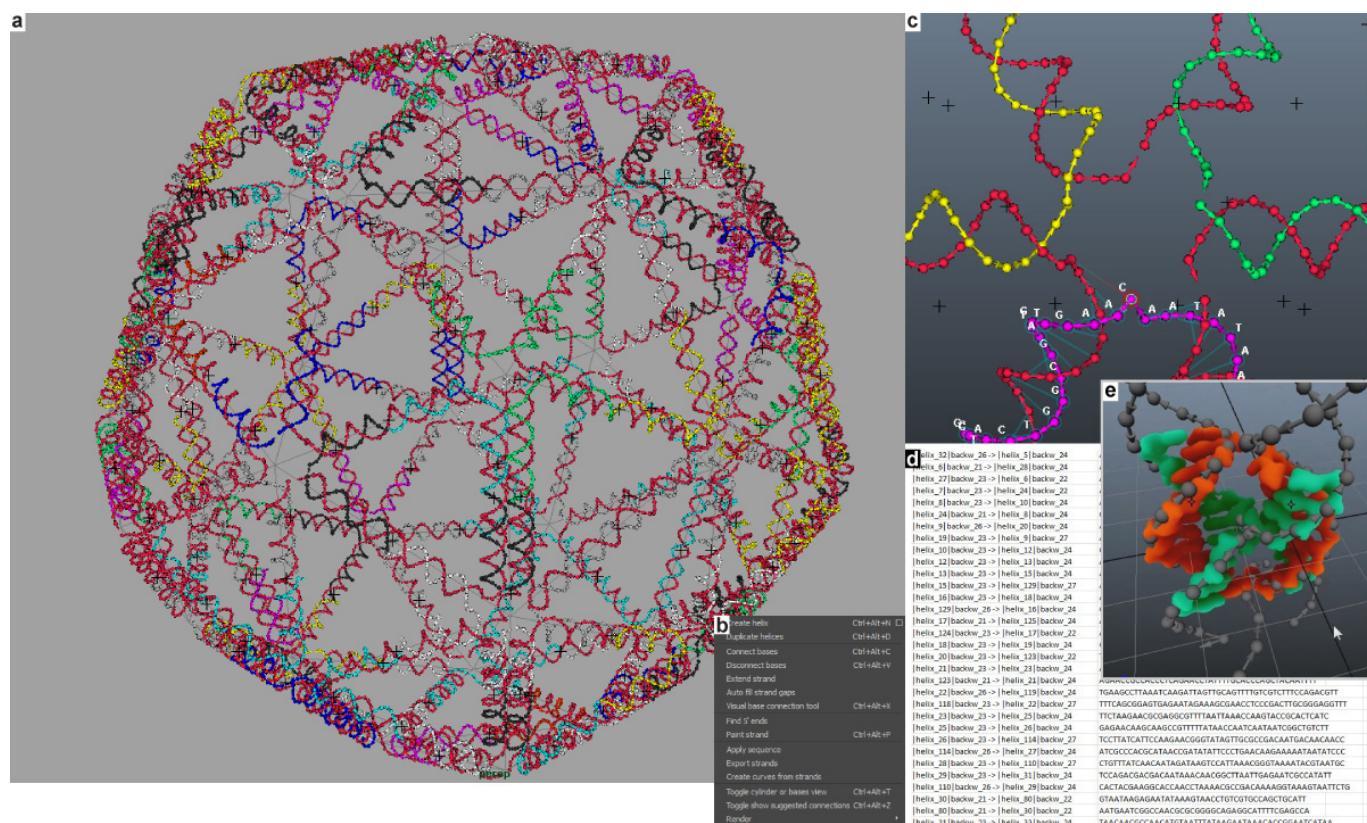


e Presented method, one edge - one helix* design goal, find routing that traverses each edge once:



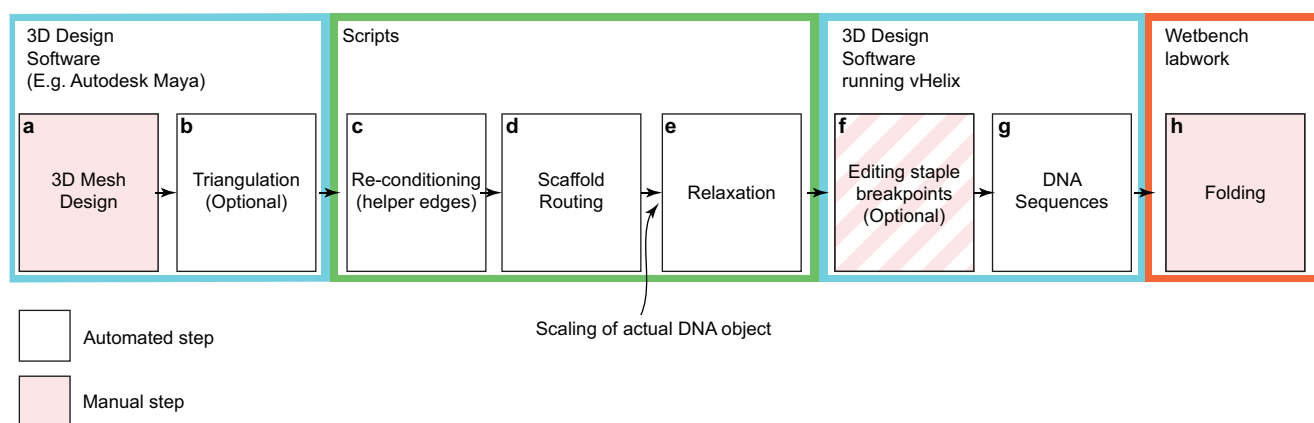
Extended Data Figure 1 | Comparison with previous strategy for polygonal DNA origami. **a**, Previous strategies for folding polygonal DNA origami have relied on folding the circular single-stranded DNA into a tree-like shape, where each branch is composed of an even number of helices (two in this illustration), these branches are then connected using helper joins as in **b**, where staple strands (in blue) bridge the gap between the distant parts of the scaffold, to yield the final polyhedral structure: the tetrahedron to the right in this example. **c**, The target shape and its flattened Schlegel representation. **d**, Previous methods have introduced helper joins in $N - 1$ of the edges, where N is the number of faces in the structure. Notably, the structures

presented in this work would require on the order of 100 helper joins. A large number of helper joins is commonly believed to increase aggregation problems owing to the sticky ends produced as intermediates during folding. **e**, The strategy presented in this work. The goal is to route the entire scaffold through all the edges of the mesh, without crossing and with preferably only one traverse per edge. *It turns out that one helix per edge is not possible for all meshes (as described in the main text, Fig. 1 and in Supplementary Note 1). Odd-degree vertices require some edges to be traversed twice by the scaffold routing.



Extended Data Figure 2 | An overview of vHelix. To be able to work with non-canonical origami designs, we implemented software that would allow free-form manipulation of helices directly in 3D space. The software was implemented as a plug-in for Autodesk Maya (several versions) and is available at <http://www.vhelix.net>. The associated source code can be found at <https://github.com/gardell/vHelix>. **a**, The interface in vHelix when viewing the design of the ball structure. **b** The 'Helix' menu provides most of the functionality, such as the ability to create new helices, disconnect and connect bases. **c**, Close-up of a connected vertex. Selecting a base shows its associated connectivity by highlighting all connected bases and displaying the

associated sequence if a sequence has been applied. **d**, Using the "apply sequence" command to one of the strands (the scaffold), the plug-in calculates the sequence of all paired bases (on the staple strands) and subsequently the command "export strands" generates a spreadsheet file containing the staple-strand sequences. The physical dimensions of the DNA model follows what is usually used in DNA nanotechnology design processes (that is, a 2-nm helical radius, a 0.334-nm rise, a 34.286° pitch and a 155° minor groove). **e**, Overlaying the model with crystallography data from the literature³⁹ shows that the model fits natural DNA well.



Extended Data Figure 3 | Design pipeline overview. **a**, We started the designs in Autodesk Maya, importing or modelling our own 3D polygon mesh object. **b**, The triangulation step is not mandatory because the scaffold routing and further processing is not limited to triangulated meshes, but it is used for all structures reported here to achieve extra rigidity by triangulation. Steps **c–e** are implemented as a series of scripts that process the mesh exported from the 3D design software. **c**, All odd-degree vertices are joined by helper edges using a minimum weight perfect matching algorithm (see Supplementary Note 1). **d**, The re-conditioned mesh is fed to a script implementing the A-trails routing algorithm (see Supplementary Note 1). **e**, After scaffold routing, the physical relaxation model reads the routed path. Up until now, the mesh has been treated as an abstract graph; in the relaxation step, however, an input is required to set the physical size of the desired DNA rendering, that is, the user sets a scaling value to fit the mesh to the scaffold available for the folding. The

relaxation simulation and length-modification scheme (described in more detail in Supplementary Note 2) will rotate and shorten/lengthen some edges to find an overall best fit to the desired 3D shape while accounting for strain between nucleotides in the vertices. The output of the relaxation/length modification optimization is a file readable by vHelix, a plug-in for Autodesk Maya. **f**, As the file is imported into vHelix, the user has the option of automatically positioning staple-strand break-points by stating parameters for maximum staple length and the minimum length of edges with breakpoints. Alternatively, the staple-strand breakpoints can be edited manually in vHelix after importing. **g**, The DNA sequences of all staple strands given a scaffold input is calculated and exported to a spreadsheet by vHelix. **h**, The mixing of staple strands and scaffold is done by hand but a pipetting robot could conceivably also make this last step highly automated.

Catalytic asymmetric umpolung reactions of imines

Yongwei Wu¹, Lin Hu¹, Zhe Li¹ & Li Deng¹

The carbon–nitrogen double bonds in imines are fundamentally important functional groups in organic chemistry. This is largely due to the fact that imines act as electrophiles towards carbon nucleophiles in reactions that form carbon–carbon bonds, thereby serving as one of the most widely used precursors for the formation of amines in both synthetic and biosynthetic settings^{1–5}. If the carbon atom of the imine could be rendered electron-rich, the imine could react as a nucleophile instead of as an electrophile. Such a reversal in the electronic characteristics of the imine functionality would facilitate the development of new chemical transformations that convert imines into amines via carbon–carbon bond-forming reactions with carbon electrophiles, thereby creating new opportunities for the efficient synthesis of amines. The development of asymmetric umpolung reactions of imines (in which the imines act as nucleophiles) remains uncharted territory, in spite of the far-reaching impact such reactions would have in organic synthesis. Here we report the discovery and development of new chiral phase-transfer catalysts that promote the highly efficient asymmetric umpolung reactions of imines with the carbon electrophile enals. These catalysts mediate the deprotonation of imines and direct the 2-azaallyl anions thus formed to react with enals in a highly chemoselective, regioselective, diastereoselective and enantioselective fashion. The reaction tolerates a broad range of imines and enals, and can be carried out in high yield with as little as 0.01 mole per cent catalyst with a moisture- and air-tolerant operational protocol. These umpolung reactions provide a conceptually new and practical approach to chiral amino compounds.

Umpolung reactions create new activities by reversing the inherent polarity of common organic functionalities such as carbonyls and consequently allow the development of new reactions of distinct bond connections⁶. The successful development of numerous C–C bond-forming umpolung reactions with carbonyls as acyl anion equivalents has greatly expanded the repertoire of organic synthesis^{7–9}. The power of carbonyl umpolung reactions has been tapped for asymmetric synthesis through the successful development of efficient chiral catalysts for enantioselective Stetter reactions and other asymmetric reactions¹⁰. In contrast, C–C bond-forming umpolung reactions of imines are

rarely reported^{11–14}. Aiming at the realization of highly efficient catalytic asymmetric umpolung reactions of imines, we embarked on a search for catalysts to both promote the formation of carbanions from imines and direct the carbanions thus formed to react with carbon electrophiles to generate chiral amines in an asymmetric fashion.

We recently reported that modified cinchona alkaloids such as the quinine-derived (Q) catalyst Q-2 could promote highly enantioselective isomerization of trifluoromethyl imines such as **1** (Fig. 1)^{15,16}. This reaction presumably proceeds through the initial formation of the 2-azaallyl anion **3**, and then a highly enantioselective protonation of **3**. This discovery prompted us to postulate that, if the 2-azaallyl anion **3** could be made to react with carbon electrophiles in a stereoselective manner, novel C–C bond-forming asymmetric reactions transforming imines **1** into enantioenriched amines could be realized (Fig. 1). Although numerous catalytic asymmetric C–C bond-forming reactions with enolates derived from glyoxylateimines¹⁴ and glycine imines¹⁷ have been documented for the synthesis of amino acids, only two catalytic asymmetric C–C bond-forming reactions with 2-azaallyl anions have been reported^{18,19}. The palladium-catalysed cross-coupling of 2-azaallyl anions with aryl halides and triflates remains the sole example of highly enantioselective C–C bond-forming reactions with 2-azaallyl anions¹⁸.

Guided by these considerations, we investigated quinine- and quinidine-derived (QD) organocatalysts Q-2, QD-11 and QD-12 for the reaction of imine **1A** and crotonaldehyde (**8a**) (Table 1). None of them was active towards the desired C–C bond-forming reaction; only the isomerized imine **4A** was detected. These catalysts promoted the deprotonation of trifluoromethyl imine **1A** to form the 2-azaallyl anion **3**, but were unable to direct the conjugate addition of **3** to crotonaldehyde. Presumably, the protonated cinchona alkaloids formed on deprotonation of **1A** rapidly protonate **3** to form **4A**. As the 2-azaallyl anion **3** was shown to engage in protonation in the presence of a proton donor, we surmise that a novel class of catalysts must be developed to afford the required chemoselectivity in favour of the C–C bond formation over the protonation.

We decided to explore chiral phase-transfer catalysts²⁰. Under phase-transfer catalysis conditions, stronger bases could be explored

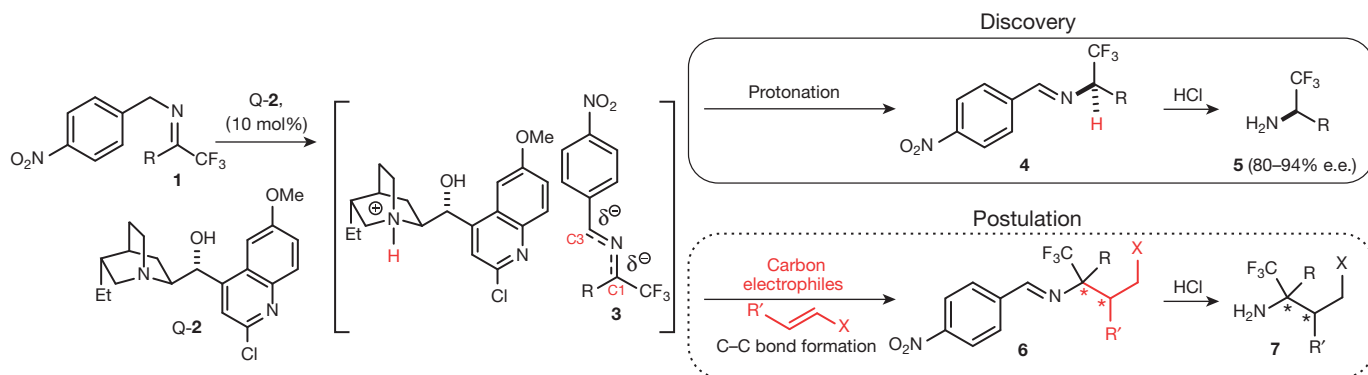
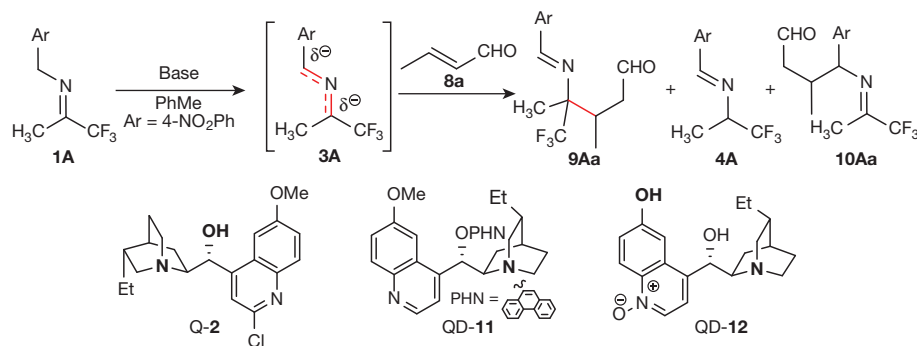


Figure 1 | Design of a catalytic C–C bond-forming umpolung reaction of imines. See text for details.

¹Department of Chemistry, Brandeis University, 415 South Street, Waltham, Massachusetts 02454, USA.

Table 1 | Experiments with chiral base catalysts



Entry	<i>T</i> (°C)	Catalyst	Conversion (%)	9/4
1	RT	Q-2	84	0/100
2	RT	QD-11	32	0/100
3	RT	QD-12	9	0/100

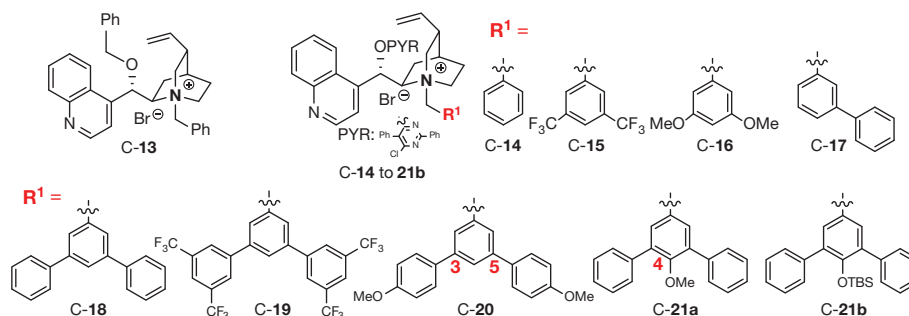
Conditions: room temperature (RT), 10 mol% catalyst, 16 h.

for the deprotonation of imine **1** to form 2-azaallyl anion **3**. Furthermore, in the absence of a protonated cationic species, **3** should be less prone to protonation and therefore more likely to engage in the addition to **8a**. A cinchonine-derived (C) phase-transfer catalyst **C-13** was first investigated to promote the reaction of **1A** and **8a** in toluene and aqueous KOH at room temperature. The desired amine **9Aa** was formed, albeit in minuscule amounts (entry 1, Table 2). Importantly, the chemoselectivity for the C–C bond formation could be improved with catalyst **C-14** bearing PYR, a bulky heteroaryl group (see Table 2 for details), although both the reaction conversion and the chemoselectivity remained poor (entry 2, Table 2). Subsequently, we found that a reaction at lower temperature afforded significantly improved conversion and chemoselectivity. The absence of **10Aa**, which would be formed by conjugate addition from the other end of the 2-azaallyl

anion, is noteworthy. However, amine **9Aa** was formed with moderate diastereoselectivity and poor enantioselectivity.

Introducing an additional interaction between a conformationally well-defined phase-transfer catalyst and the anionic nucleophile has proven to be a useful strategy to enhance catalytic selectivity²¹. We hypothesized that a cinchonine-derived phase-transfer catalyst bearing a properly located aromatic group with suitable electronic properties might interact with 2-azaallyl anion **3A** via both ionic and π – π interactions^{22–24}, thereby mediating the model umpolung reaction in a highly chemo-, regio-, diastereo- and enantioselective fashion. Analogues **C-15** and **C-16** bearing electron-withdrawing and electron-donating *N*-benzyl substituents, respectively, were examined. We found that **C-16** afforded only improved conversion whereas **C-15** was worse than **C-14** (entries 4 and 5, Table 2). Interestingly,

Table 2 | Screening and optimization of chiral phase-transfer catalysts



Entry	<i>T</i> (°C)	Catalyst	Conversion (%)	9/4; 9/10	d.r. of 9	e.e. of 9 (%)
1	RT	C-13	41	2/98; ND	ND	ND
2	RT	C-14	18	11/89; ND	ND	ND
3	–20	C-14	58	37/63; >95/5	82/18	39
4	–20	C-15	54	36/64; >95/5	67/33	18
5	–20	C-16	84	34/66; >95/5	76/24	40
6*	–20	C-16	41	32/68; >95/5	74/26	39
7*	–20	C-17	14	67/33; >95/5	87/13	68
8*	–20	C-18	40	74/26; >95/5	86/14	77
9*	–20	C-19	39	45/55; >95/5	96/4	55
10*	–20	C-20	66	68/32; >95/5	91/9	85
11*	–20	C-21a	88	94/6; >95/5	91/9	91
12*	–20	C-21b	99	99/1; >95/5	93/7	96
13†	–20	C-21b	97	99/1; >95/5	93/7	95
14*	–20	TBAB	31	4/96; ND	ND	ND

Conditions: 10 mol% catalyst, 10 mol% KOH(aq), 16 h. TBAB, tetra-*n*-butylammonium bromide. ND, not determined; d.r., diastereomeric ratio; e.e., enantiomeric ratio.

*1.0 mol% catalyst, 10 mol% KOH(aq), 2 h.

†0.2 mol% of **C-21b** used, 5 h.

we observed that a decrease in the loading of C-16 did not affect the catalytic selectivities negatively (entry 6 versus 5, Table 2). We therefore decreased the catalyst loading from 10 mol% to 1 mol% in our subsequent catalyst screening and optimization studies. We next turned to C-17, an analogue containing a biphenyl group. C-17 afforded substantially improved chemo-, diastereo- and enantioselectivity, thereby allowing amine **9Aa** to be formed as the major product (entry 7 versus 6, Table 2).

Assuming the improved catalysis resulted from a π - π interaction between the biaryl moiety of C-17 and **3A**, we designed and synthesized catalyst C-18 (Table 2). We reasoned that the presence of the C2-symmetric terphenyl moiety could render C-18 a more efficient catalyst than C-17. This working hypothesis received support from the superior performance of C-18 in catalytic activity as well as chemo- and enantioselectivity (entry 8 versus 7, Table 2). Further tuning of the terphenyl moiety was initially attempted by introducing electron-withdrawing and electron-donating groups on the 3- and 5-phenyl groups. Catalyst C-20 (entry 10, Table 2) bearing an electron-rich terphenyl group performed better than C-19 (entry 9), which contained an electron-deficient terphenyl moiety. However, C-20 furnished higher stereoselectivity but lower chemoselectivity than those produced by C-18 (entry 10 versus 8, Table 2).

We next examined catalyst C-21a (Table 2) which was designed to create an electron-rich terphenyl moiety with an electron-donating

substituent in a position not causing obstructive steric interference between the catalyst and 2-azaallyl anion **3**. Gratifyingly, C-21a not only turned out to be much more active, but also afforded **9Aa** with synthetically useful chemo-, regio-, diastereo- and enantioselectivity (entry 11, Table 2). Catalyst C-21b with a more electron-donating and bulky *tert*-butyldimethylsilyl ether (OTBS) group was more active and selective; a loading of only 0.2 mol% produced imine **9Aa** rapidly with almost complete chemoselectivity and excellent stereoselectivity (entry 13, Table 2). We attributed the superiority of C-21b over C-21a to two factors resulting from the substitution of the 4-methoxy with the 4-OTBS group: (1) the terphenyl moiety is more electron rich due to the presence of the more electron-donating 4-OTBS group; (2) the terphenyl moiety has less conformational flexibility due to steric hindrance of the rotation of the 3,5-phenyl rings by the bulky 4-OTBS group. Both factors could reinforce the π - π interaction between **3A** and the catalyst C-21b.

Only a trace of **9Aa** was formed from **1A** and **8a** using tetrabutylammonium bromide (TBAB) as the quaternary ammonium salt (entry 14, Table 2), which confirmed that the structural characteristics of C-21b were responsible for both the catalytic activity and the selectivity observed for the umpolung reaction between imine **1A** and enal **8a**. To ascertain that 2-azaallyl anion **3** originated only from imine **1** rather than also from the isomerized imine **4**, we established that no reaction occurred between **4A** and **8a** under the optimized conditions.

Table 3 | Substrate scope for umpolung reactions of trifluoromethyl imines with enals

Entry	R ¹	Time (h); conversion (%)	9/4; 9/10	d.r. of 9	Yield (%) [*]	e.e. (%) [†]
1	H ₃ C- 1A	5; 99	>95/5; >95/5	93/7	81 (22Aa)	95
2	1B	5; 97	>95/5; >95/5	91/9	84 (22Ba)	94
3	1C	5; 98	>95/5; >95/5	91/9	83 (22Ca)	96
4	Br- 1D	5; 99	>95/5; >95/5	91/9	75 (22Da)	96
5	BnO- 1E	7; 94	>95/5; >95/5	91/9	72 (22Ea)	96
6	Cy- 1F	12; 98	91/9; >95/5	93/7	54 (22Fa)	95

Scope of β -substituted enals in reactions with imine **1A**

Entry	R ²	Time (h); conversion (%)	9/4; 9/10	d.r. of 9	Yield (%) [*]	e.e. (%) [†]
7	CH ₃ CH ₂ ; 8b	5; 99	89/11; >95/5	>95/5	64 (22Ab)	95
8	CH ₃ (CH ₂) ₅ ; 8c	12; 93	86/14; >95/5	>95/5	51 (22Ac)	96
9	Ph; 8d	8; 93	>95/5; 68/32	>95/5	51 (22Ad)	91

Scope of imines in reactions with acrolein (8e**, R² = H)**

Entry	R ¹	Time (h); conversion (%)	9/4; 9/10	Yield (%) [*]	e.e. (%) [†]
10 [‡]	H ₃ C- 1A	3; 95	>95/5; >95/5	89 (22Ae)	92
11 [‡]	1B	3; 99	>95/5; >95/5	82 (22Be)	91
12 [‡]	Br- 1D	3; 97	>95/5; >95/5	84 (22De)	91
13 [‡]	Cy- 1F	3; 99	>95/5; >95/5	90 (22Fe)	92
14	Ph; 1G	3; 99	94/6; >95/5	71 (23Ge)	94
15	<i>p</i> -MeOC ₆ H ₄ ; 1H	3; 94	92/8; >95/5	67 (23He)	94
16	<i>p</i> -CF ₃ C ₆ H ₄ ; 1I	3; 99	88/12; >95/5	78 (23Ie)	92
17	Ph- 1J	1; 99	>95/5; >95/5	90 (23Je)	93

Conditions: imine **1** (0.2 mmol), aldehyde **8** (0.4 mmol), C-21b (0.2 mol%), KOH (2.2 μ l, 50 wt% aq., 10 mol%), PhMe (2.0 ml). Conversion, regioselectivity (9/10) and d.r. of **9** were determined by ¹H NMR analysis of the crude umpolung reaction mixture. Chemoselectivity (9/4) was determined by ¹⁹F NMR analysis.

^{*}Overall yield for the transformation of imine **1** to either **22** or **23**.

[†]e.e. of **22** or **23**, determined by HPLC analysis.

[‡]Reaction was performed at -10 °C.

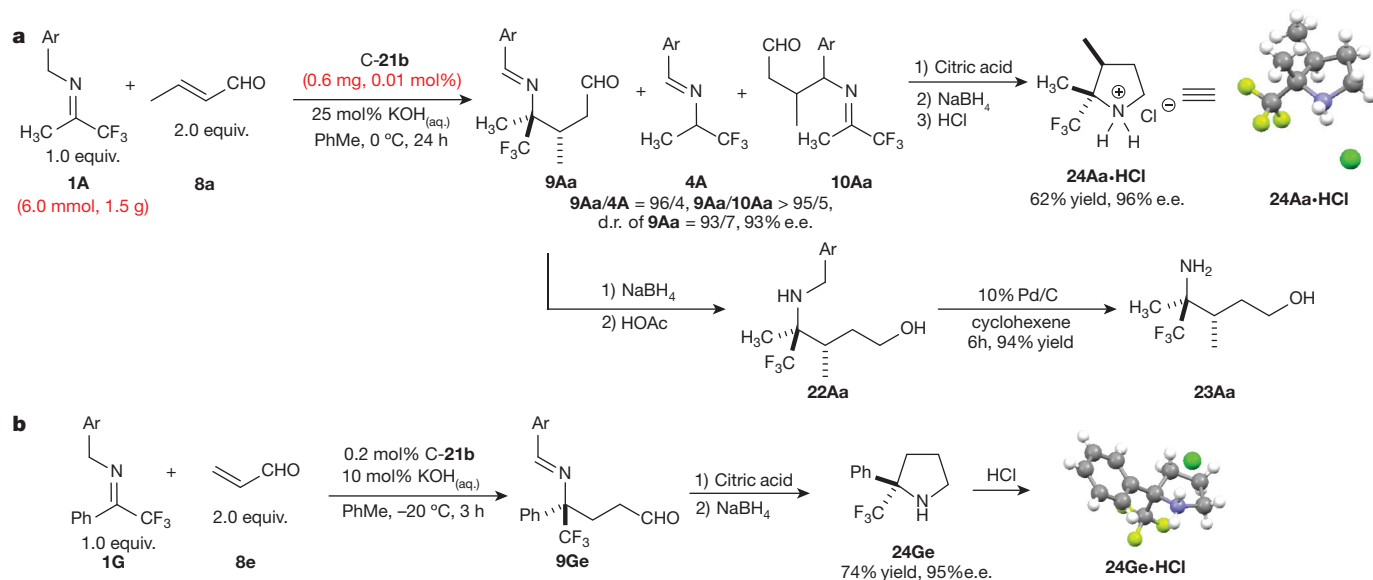


Figure 2 | Gram-scale reaction and synthetic applications. **a**, Highly efficient gram-scale catalytic asymmetric umpolung reaction of imine **1A** (1.5 g) with 0.01 mol% of **C-21b** (0.6 mg) and its application for the syntheses of aminoalcohol **23Aa** and pyrrolidine **24Aa**. **b**, Synthetic application of catalytic asymmetric umpolung reaction of imine **1G** for the syntheses of amine **9Ge** and pyrrolidine **24Ge**.

It should be noted that amine **9Aa** may also form via a [3+2] cycloaddition between **1A** and **8a** followed by a retro-Mannich reaction. However, we did not detect the formation of the [3+2] adduct when monitoring the reaction by ^1H and ^{19}F NMR analyses.

Our investigation of the substrate scope began with the reaction of **1A** and **8a** with 0.2 mol% of **C-21b** (entry 1, Table 3). The reaction proceeded to full conversion within 5 h with excellent chemo-, regio-, diastereo- and enantioselectivities. The optically active amine **9Aa** was then converted to the more stable *N*-benzyl aminoalcohol **22Aa** by reducing first the aldehyde with NaBH_4 and then the imine with NaBH_4 and acetic acid, which could be readily isolated as a single diastereomer in good yield. Reactions of **8a** with a series of trifluoromethyl imines (**1B–E**, Table 3) bearing simple and functionalized linear alkyl substituents consistently proceeded in high yield and excellent chemoselectivity and stereoselectivity. The reaction tolerated an imine bearing a β -branched alkyl substituent (**1F**). The reaction accepted larger β -alkyl groups on the enal (entries 7 and 8, Table 3). Cinnamaldehyde (**8d**) reacted with **1A** to give a 68:32 mixture of the

desirable amine **9Ad** and the regioisomer **10Ad**. Nonetheless, **9Ad** was produced with high chemo-, diastereo- and enantioselectivity in synthetically useful yield (entry 9, Table 3).

We next examined the reactions of trifluoromethylated imines **1** with acrolein (**8e**). We found that at -10°C the reaction between **1A** and **8e** proceeded cleanly and in a highly enantioselective fashion to furnish the corresponding amine **9Ae** as the only detectable product by NMR analysis of the crude reaction mixture. The reactions of acrolein (**8e**) with trifluoromethyl imines **1** bearing a variety of alkyl, aryl and alkenyl substituents were equally successful, affording the corresponding trifluoromethylated amines **9** containing a tetrasubstituted stereocentre^{25,26} in high optical purity (entries 11–17, Table 3). Alkyl trifluoromethylated amines (**9Ae–Fe**) were converted to *N*-benzyl aminoalcohols **22** (entries 10–13, Table 3). Aryl and alkenyl amines **9Ge–Je** were converted to aminoalcohols **23** by reduction of the aldehyde with NaBH_4 and hydrolysis of the imine with aqueous HCl (entries 14–17, Table 3). In all these cases, the aminoalcohols **22** and **23** were obtained in good yields and high optical purity.

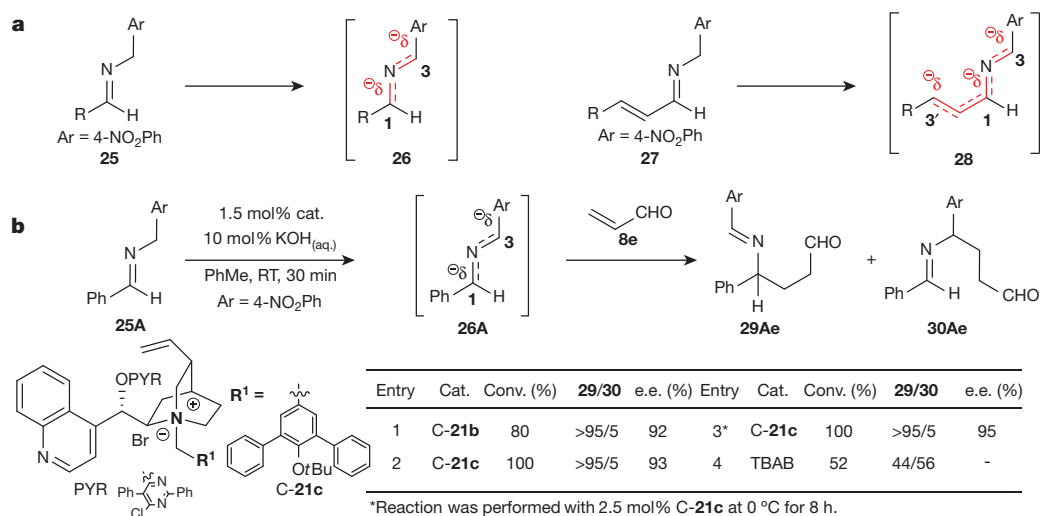


Figure 3 | Asymmetric umpolung reactions of aryl and unsaturated aldimines. **a**, Left, 2-azaallyl anion **26** derived from deprotonation of phenyl imine **25**; right, 2-azaallyl anion **28** derived from deprotonation of alkenyl imine **27**. **b**, Catalyst optimization for the umpolung reaction of phenyl imine **25A** with enal **8e**.

Table 4 | Substrate scope for umpolung reactions of aryl aldimines with acrolein (8e**)**

Entry	R	Time (h)	29/30	Yield of 31 (%) [*]	e.e. of 31 (%) [†]
1	Ph; 25A	8	>95/5	55	93
2	<i>o</i> -CH ₃ C ₆ H ₄ ; 25B	8	>95/5	51	94
3 [‡]	2-Naphthyl; 25C	8	90/10	54	94
4	2-Thienyl; 25D	8	>95/5	53	95
5 [‡]	<i>p</i> -BrC ₆ H ₄ ; 25E	5	>95/5	52	95
6	<i>o</i> -BrC ₆ H ₄ ; 25F	5	>95/5	56	95
7 [‡]	<i>p</i> -MeO ₂ CC ₆ H ₄ ; 25G	8	83/17	53	90
8 [§]	<i>p</i> -MeOC ₆ H ₄ ; 25H	18	>95/5	45	95

Conditions: reactions were performed with **25** (0.20 mmol), **8e** (0.40 mmol), C-**21c** (2.5 mol%) and KOH (2.2 μ l, 50 wt% aq., 10 mol%) in PhMe (2.0 ml) until full conversion. Regioselectivity (**29/30**) was determined by ¹H analysis of the crude umpolung reaction mixture.

^{*}Overall yield for the transformation of imine **25** to **31**.

[†]Determined by HPLC analysis.

[‡]Reaction was performed in PhMe/CH₂Cl₂ = 2/1 solution (3.0 ml).

[§]5.0 mol% C-**21c** used.

A gram-scale reaction of **1A** with **8a** with 0.01 mol% of C-**21b** went to completion without deterioration in selectivity (Fig. 2a). This remarkable catalytic efficiency indicates the utility of this new reaction in preparative-scale organic synthesis²⁷. To demonstrate the synthetic versatility of this reaction, we converted chiral aminoaldehyde **9Aa** to aminoalcohol **23Aa** and pyrrolidine **24Aa** as shown in Fig. 2a. Similarly, the phenyl substituted product **9Ge** was converted to pyrrolidine **24Ge** (Fig. 2b). The absolute configurations of **24Aa** and **24Ge** were determined by X-ray crystallography.

We are interested in extending the scope to simple imines, which would greatly expand the reach of this asymmetric umpolung reaction in organic synthesis. However, 2-azaallyl anions **26** derived from aryl imines **25** (Fig. 3a) are substantially less stable than those derived from the corresponding trifluoromethyl imines **1**. Furthermore, regioselectivity control for the electrophilic reaction with an unsymmetrically substituted 1,3-diaryl-2-azaallyl anion **26** might prove difficult (Fig. 3a). For example, deprotonation of phenyl imine **25A** should form

2-azaallyl anion **26A**, which is flanked by the phenyl and the 4-nitrophenyl rings (Fig. 3b). Thus, there is an inherent electronic bias for an electrophile to react with **26A** by attacking preferentially the more electron-rich C3²⁸. Nonetheless, the remarkable catalytic efficiency of C-**21b** made us hopeful that it could provide powerful catalytic activity and selectivity to overcome this undesirable substrate bias while still affording the required stereoselectivity for an efficient asymmetric imine umpolung reaction.

Accordingly, we investigated the reaction of phenyl imine **25A** with acrolein (**8e**) applying the conditions established with trifluoromethyl imines **1**. As expected, **25A** was far less reactive than **1A**; only a trace amount of the desired product **29Ae** was detected. With a substantially increased catalyst loading (entry 1, Fig. 3b), the reaction progressed to high conversion and in excellent enantioselectivity. A new catalyst bearing a 4-*o*Tu group (C-**21c**) was found to be more active and afforded better enantioselectivity (entry 2, Fig. 3b); this allowed a clean and complete reaction to occur at 0 °C in excellent enantioselectivity

Table 5 | Substrate scope for umpolung reactions of alkenyl aldimines with acrolein (8e**)**

Entry	Alkenyl	Time (h)	32/33	Yield of 34 (%) [*]	e.e. of 34 (%) [†]
1	Ph-CH=CH-; 27A	16	86/14	51	92
2	Me-CH=CH-; 27B	16	95/5	50	92
3 [‡]	Ph-CH=CH-; 27C	24	82/18	46	95
4	<i>p</i> -BrC ₆ H ₄ -CH=CH-; 27D	12	77/23	44	92
5 [‡]	<i>p</i> -MeOC ₆ H ₄ -CH=CH-; 27E	24	83/17	41	90
6 [‡]	Me-CH=CH-; 27F	6	95/5	37 [§]	90

Conditions: reactions were performed with **27** (0.20 mmol), **8e** (0.40 mmol), C-**21c** (2.5 mol%) and KOH (2.2 μ l, 50 wt% aq., 10 mol%) in PhMe (2.0 ml) until full conversion. Regioselectivity (**32/33**) was determined by ¹H analysis of the crude umpolung reaction mixture.

^{*}Overall yield for the transformation of imine **27** to **34**.

[†]Determined by HPLC analysis.

[‡]5.0 mol% C-**21c** used.

[§]Overall yield for a four-step transformation of (*E*)-3-bromobut-2-enal to **34Fe**, see Supplementary Information for details.

with 2.5 mol % of C-21c (entry 3, Fig. 3b). Amine 29Ae was converted to the Boc-protected aminoalcohol 31Ae in high optical purity and good yield in three steps (entry 1, Table 4). Subsequently, we established that the umpolung reaction tolerated a broad range of aryl and heteroaryl aldimines of varying steric and electronic properties (entries 2–8, Table 4). Electron-rich aryl imines such as 25H appeared to be less active, but the umpolung reaction with C-21c still went to completion with high chemoselectivity, regioselectivity and enantioselectivity.

Owing to the synthetic versatility of the olefin and amine functionalities, chiral allylic amines are highly valuable chiral building blocks²⁹. If we could extend the substrate scope to α,β -unsaturated imines 27 (Fig. 3a), the impact of the imine umpolung reactions would be further enlarged. However, the 2-azaallyl anions 28 derived from α,β -unsaturated imines 27 were expected to be even less stable than those derived from arylaldehydes³⁰. Furthermore, the conjugation of an azaallyl anion with an olefin renders 28 a more challenging nucleophile from the viewpoint of achieving catalytic control of regioselectivity (Fig. 3a). Gratifyingly, C-21c provided highly selective catalysis to efficiently promote the umpolung reaction of 27A and 8e (entry 1, Table 5). Importantly, the efficiency of C-21c remained undiminished for reactions involving a variety of α,β -unsaturated imines bearing di- and trisubstituted olefins (entries 2–6, Table 5). As allylic amines could be readily hydrogenated to the corresponding aliphatic amines (Table 5), these results established this imine umpolung reaction as a useful method for the asymmetric synthesis of both chiral allylic and aliphatic amines.

We have identified a new class of tunable chiral phase-transfer catalysts and demonstrated their unique ability to promote C–C bond-forming reactions with 2-azaallyl anions in a highly chemoselective, regioselective, diastereoselective and enantioselective fashion. This discovery releases the potential of imines as nucleophiles, thereby allowing the realization of catalytic asymmetric umpolung reactions of imines, and providing a fundamentally new approach towards chiral amino compounds. With a simple operational protocol and low catalyst loading, this transformation also provides a practical method for organic synthesis.

Received 23 February; accepted 21 May 2015.

- Nugent, T. C. *Chiral Amine Synthesis: Methods, Developments and Applications* (Wiley-VCH, 2010).
- Robak, M. T., Herbage, M. A. & Ellman, J. A. Synthesis and applications of tert-butanefulfonamide. *Chem. Rev.* **110**, 3600–3740 (2010).
- Kobayashi, S., Mori, Y., Fossey, J. S. & Salter, M. M. Catalytic enantioselective formation of C–C bonds by addition to imines and hydrazones: a ten-year update. *Chem. Rev.* **111**, 2626–2704 (2011).
- Silverio, D. L. *et al.* Simple organic molecules as catalysts for enantioselective synthesis of amines and alcohols. *Nature* **494**, 216–221 (2013).
- Dewick, P. M. *Medicinal Natural Products: A Biosynthetic Approach* 3rd edn (Wiley & Sons, 2009).
- Seebach, D. Methods of reactivity umpolung. *Angew. Chem. Int. Ed. Engl.* **18**, 239–258 (1979).
- Seebach, D. & Corey, E. J. Generation and synthetic applications of 2-lithio-1,3-dithianes. *J. Org. Chem.* **40**, 231–237 (1975).
- Smith, A. B. & Adams, C. M. Evolution of dithiane-based strategies for the construction of architecturally complex natural products. *Acc. Chem. Res.* **37**, 365–377 (2004).
- Brehme, R., Enders, D., Fernandez, R. & Lassaletta, J. M. Aldehyde N-dialkylhydrazones as neutral acyl anion equivalents: umpolung of the imine reactivity. *Eur. J. Org. Chem.* 5629–5660 (2007).
- Vora, H. U. & Rovis, T. Asymmetric N-heterocyclic carbene (NHC) catalyzed acyl anion reactions. *Aldrichim. Acta* **44**, 3–11 (2011).
- Reich, B. J. E., Justice, A. K., Beckstead, B. T., Reibenspies, J. H. & Miller, S. A. Cyanide-catalyzed cyclizations via aldimine coupling. *J. Org. Chem.* **69**, 1357–1359 (2004).
- Ogle, J. W., Zhang, J., Reibenspies, J. H., Abboud, K. A. & Miller, S. A. Synthesis of electronically diverse tetraarylimidazolidene carbenes via catalytic aldimine coupling. *Org. Lett.* **10**, 3677–3680 (2008).
- Liu, X., Gao, A., Ding, L., Xu, J. & Zhao, B. Aminative umpolung synthesis of aryl vicinal diamines from aromatic aldehydes. *Org. Lett.* **16**, 2118–2121 (2014).
- Matsumoto, M., Harada, M., Yamashita, Y. & Kobayashi, S. Catalytic imine-imine cross-coupling reactions. *Chem. Commun.* **50**, 13041–13044 (2014).
- Wu, Y. & Deng, L. Asymmetric synthesis of trifluoromethylated amines via catalytic enantioselective isomerization of imines. *J. Am. Chem. Soc.* **134**, 14334–14337 (2012).
- Liu, M., Li, J., Xiao, X., Xie, Y. & Shi, Y. An efficient synthesis of optically active trifluoromethyl aldimines via asymmetric biomimetic transamination. *Chem. Commun.* **49**, 1404–1406 (2013).
- Jakubowska, A. & Kulig, K. Progress in the glycine equivalent based α -amino acids synthesis. *Curr. Org. Synth.* **10**, 547–563 (2013).
- Zhu, Y. & Buchwald, S. L. Ligand-controlled asymmetric arylation of aliphatic α -amino anion equivalents. *J. Am. Chem. Soc.* **136**, 4500–4503 (2014).
- Qian, X. *et al.* Palladium-catalyzed decarboxylative generation and asymmetric allylation of α -imino anions. *Org. Lett.* **16**, 5228–5231 (2014).
- Shirakawa, S. & Maruoka, K. Recent developments in asymmetric phase-transfer reactions. *Angew. Chem. Int. Ed. Engl.* **52**, 4312–4348 (2013).
- Ooi, T., Ohara, D., Tamura, M. & Maruoka, K. Design of new chiral phase-transfer catalysts with dual functions for highly enantioselective epoxidation of α,β -unsaturated ketones. *J. Am. Chem. Soc.* **126**, 6844–6845 (2004).
- Dolling, U. H., Davis, P. & Grabowski, E. J. J. Efficient catalytic asymmetric alkylations. 1. Enantioselective synthesis of (+)-indacrinone via chiral phase-transfer catalysis. *J. Am. Chem. Soc.* **106**, 446–447 (1984).
- Bandini, M., Bottoni, A., Eichholzer, A., Miscione, G. P. & Stenta, M. Asymmetric phase-transfer-catalyzed intramolecular N-alkylation of indoles and pyrroles: a combined experimental and theoretical investigation. *Chem. Eur. J.* **16**, 12462–12473 (2010).
- Knowles, R. R., Lin, S. & Jacobsen, E. N. Enantioselective thiourea-catalyzed cationic polycyclizations. *J. Am. Chem. Soc.* **132**, 5030–5032 (2010).
- Fu, P., Snapper, M. L. & Hoveyda, A. H. Catalytic asymmetric alkylations of ketimines. Enantioselective synthesis of N-substituted quaternary carbon stereogenic centers by Zr-catalyzed additions of dialkylzinc reagents to aryl-, alkyl-, and trifluoroalkyl-substituted ketimines. *J. Am. Chem. Soc.* **130**, 5530–5541 (2008).
- Nie, J., Guo, H.-C., Cahard, D. & Ma, J.-A. Asymmetric construction of stereogenic carbon centers featuring a trifluoromethyl group from prochiral trifluoromethylated substrates. *Chem. Rev.* **111**, 455–529 (2011).
- Giacalone, F., Gruttadauria, M., Agrigento, P. & Noto, R. Low-loading asymmetric organocatalysis. *Chem. Soc. Rev.* **41**, 2406–2447 (2012).
- Bordwell, F. G., Algrim, D. & Vanier, N. R. Acidities of anilines and toluenes. *J. Org. Chem.* **42**, 1817–1819 (1977).
- Hartwig, J. F. & Stanley, L. M. Mechanistically driven development of iridium catalysts for asymmetric allylic substitution. *Acc. Chem. Res.* **43**, 1461–1475 (2010).
- Jaun, B., Schwarz, J. & Breslow, R. Determination of the basicities of benzyl, allyl, and tert-butylpropargyl anions by anodic oxidation of organolithium compounds. *J. Am. Chem. Soc.* **102**, 5741–5748 (1980).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful for financial support from the National Institute of General Medical Science (NIH, GM-61591). We thank M. Bezpalko and B. Foxman for X-ray crystallographic characterizations of structures. C. Fei and B. Hu are acknowledged for the help in substrate preparation.

Author Contributions Y.W., L.H. and Z.L. performed the experiments and analysed data. Y.W. and L.D. conceived the idea and prepared this manuscript with feedback from L.H. and Z.L.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.D. (deng@brandeis.edu).

Statistical analysis of iron geochemical data suggests limited late Proterozoic oxygenation

Erik A. Sperling^{1,2†}, Charles J. Wolock³, Alex S. Morgan¹, Benjamin C. Gill⁴, Marcus Kunzmann⁵, Galen P. Halverson⁵, Francis A. Macdonald¹, Andrew H. Knoll^{1,3} & David T. Johnston¹

Sedimentary rocks deposited across the Proterozoic–Phanerozoic transition record extreme climate fluctuations, a potential rise in atmospheric oxygen or re-organization of the seafloor redox landscape, and the initial diversification of animals^{1,2}. It is widely assumed that the inferred redox change facilitated the observed trends in biodiversity. Establishing this palaeoenvironmental context, however, requires that changes in marine redox structure be tracked by means of geochemical proxies and translated into estimates of atmospheric oxygen. Iron-based proxies are among the most effective tools for tracking the redox chemistry of ancient oceans^{3,4}. These proxies are inherently local, but have global implications when analysed collectively and statistically. Here we analyse about 4,700 iron-speciation measurements from shales 2,300 to 360 million years old. Our statistical analyses suggest that subsurface water masses in mid-Proterozoic oceans were predominantly anoxic and ferruginous (depleted in dissolved oxygen and iron-bearing), but with a tendency towards euxinia (sulfide-bearing) that is not observed in the Neoproterozoic era. Analyses further indicate that early animals did not experience appreciable benthic sulfide stress. Finally, unlike proxies based on redox-sensitive trace-metal abundances^{1,5,6}, iron geochemical data do not show a statistically significant change in oxygen content through the Ediacaran and Cambrian periods, sharply constraining the magnitude of the end-Proterozoic oxygen increase. Indeed, this re-analysis of trace-metal data is consistent with oxygenation continuing well into the Palaeozoic era. Therefore, if changing redox conditions facilitated animal diversification, it did so through a limited rise in oxygen past critical functional and ecological thresholds, as is seen in modern oxygen minimum zone benthic animal communities^{7–9}.

Proxies such as iron-speciation chemistry record the redox state of local water masses immediately above accumulating sediments. Decades of work on the behaviour of iron in marine sediments underpin the observation that enrichments in total (Fe_{tot}) and highly reactive (Fe_{hr}) iron phases track water-column redox conditions (Fe_{hr} refers to iron in pyrite plus iron that is reactive to sulfide on early diagenetic timescales)^{3,4}. This robust calibration permits the differentiation between oxic and anoxic water columns, as well as whether anoxic waters were iron- or sulfide-bearing (this calculation is based on the proportion of highly reactive iron that has been converted to pyrite, Fe_{pyr}).

Early studies of iron speciation in Proterozoic shales supported the prediction¹⁰ of euxinia in subsurface waters of Mesoproterozoic oceans and further suggested deep-ocean oxygenation late in the Neoproterozoic era^{11,12}. However, and perhaps not surprisingly, a more complex and heterogeneous pattern of Earth surface evolution emerged as additional studies increased temporal and spatial coverage. For example, marine strata deposited about 1,500 million years (Myr) ago from different localities show evidence of euxinic, ferruginous and oxic basins^{11,13,14}. Similarly, Ediacaran deep-water sediments in

Newfoundland indicate oxygenation at 580 Myr ago¹², yet coeval deep-water deposits in the Canadian Cordillera show an increasing prevalence of anoxia¹⁵, or no change at all¹⁶. Such regional heterogeneity is expected given local controls on water-column redox, and highlights the fact that iron-speciation analyses of a single section or basin cannot be extrapolated to the global ocean.

Palaeontologists have long contended with an analogous problem: how to infer global diversity through time from fossil assemblages in local stratigraphic sections. The solution was to treat tabulated data within a global statistical framework¹⁷. Following this template, we have developed a data set of about 4,700 new and published iron-speciation measurements from fine-grained clastic rocks with which to test hypotheses of global redox change in Proterozoic/Palaeozoic oceans and the potential links to animal evolution. Importantly, local proxy data in a global framework can track both the mean and variance of palaeoenvironmental conditions through time. In addition to compiling data spanning the Great Oxidation Event (GOE, around 2,300 Myr ago) through the end-Devonian period, we provide 842 new analyses from Russia, northwestern Canada, Mongolia, Namibia, Svalbard, East Greenland and the western United States (Supplementary Table 2), focusing on Neoproterozoic and Cambrian strata.

Time-binned analysis of the entire data set begins with the most basic distinctions: geographic region and depositional environment (inner shelf, outer shelf, and basinal; following refs 11, 15). We note that the basinal environment does not represent true deep-ocean depths in a modern oceanographic sense, but rather the deepest environments represented by sediments deposited during maximum flooding; ‘basinal’ therefore refers to a recognizable and consistent sub-wave base environment that has been used to track deeper-water redox conditions through time (see Supplementary Information). To test for statistically significant differences, data were compared using analysis of variance (ANOVA) and Kruskal–Wallis tests depending on normality of the data. Post-hoc Tukey–Kramer tests ($\alpha = 0.05$), pairwise Wilcoxon tests and Steel–Dwass tests were applied to explore significant differences between time bins (see Supplementary Information for binning rationale and sensitivity analyses).

We first investigated the proportion of anoxic water columns through time. It has been hypothesized that a major oxygenation event occurred around the Proterozoic–Phanerozoic transition, oxygenating the world’s deep oceans and facilitating Cambrian animal diversification. This idea has been bolstered by redox-sensitive trace-metal abundance data, which show evidence of increasing oxygen levels^{1,5,6}, although the timing and magnitude remain poorly resolved^{1,2}. Aggregated iron-speciation data provide an informative complement to global trace-metal data. Since the redox state of basinal water masses has traditionally been used as a proxy for the overall ocean–atmosphere system, and shallow-water samples are rare and heterogeneously distributed through time (Supplementary Table 1), this

¹Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA. ²Integrative Oceanography Division, Scripps Institution of Oceanography, La Jolla, California 92037, USA. ³Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ⁴Department of Geosciences, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, USA. ⁵Department of Earth and Planetary Sciences/GEOTOP, McGill University, Montreal, Quebec, H3A 0E8, Canada. [†]Present address: Department of Geological and Environmental Sciences, Stanford University, Stanford, California 94305, USA.

analysis includes only samples from outer shelf and basinal environments. The proportion of samples probably deposited beneath an anoxic water column ($\text{Fe}_{\text{hr}}/\text{Fe}_{\text{tot}} > 0.38$)³ was calculated for each region, and the mean and standard error were determined for each time bin. In contrast to trace-metal data, analysis of iron-speciation data does not show a significant change in the proportion of anoxic water columns from the Proterozoic into the early Palaeozoic (ANOVA $F_{4,52} = 0.78$, $P = 0.54$; Kruskal–Wallis $\chi^2 = 3.30$, $P = 0.51$) (Fig. 1a and Supplementary Table 4), which is consistent with qualitative observations in a previous compilation¹⁵.

Iron speciation more robustly identifies anoxia as opposed to oxic conditions, because Fe_{hr} enrichments can be muted during rapid deposition or in pervasively anoxic oceans where mass-balance requirements may not result in modern-like iron enrichment. Nonetheless, the proportion of oxic samples (using a conservative threshold of $\text{Fe}_{\text{hr}}/\text{Fe}_{\text{tot}} < 0.22$)³ was tested, and again no significant differences were found (Supplementary Table 4). This result raises a number of questions that we discuss below, ranging from diagnosing the nature of basinal anoxia to reconciling the seemingly divergent results between trace-metal geochemistry and our database analysis.

To assess the nature of anoxic waters through time we focused on samples from deeper-water environments with $\text{Fe}_{\text{hr}}/\text{Fe}_{\text{tot}} > 0.38$. The average proportion of ferruginous samples between 2,300 Myr ago and 1,000 Myr ago is 0.59 (the balance being euxinic), consistent with recent arguments that basinal waters through the middle of the Proterozoic were predominantly ferruginous^{3,13} (the effect of subdividing the Proterozoic using a shorter time bin of 1,600–1,000 Myr ago was also tested; Supplementary Table 4). In fact, anoxic waters throughout the Proterozoic and Palaeozoic are more likely to be ferruginous than euxinic. However, real differences exist between time bins (Kruskal–Wallis $\chi^2 = 13.9$, $P = 0.008$). Specifically, the late Palaeoproterozoic/Mesoproterozoic bin is more likely to capture euxinic conditions than the early Neoproterozoic, Ediacaran and Cambrian intervals, where the proportion of ferruginous samples approaches unity. The Ordovician–Devonian then marks a return to limited euxinia that is statistically distinct from the Neoproterozoic bins (Fig. 1b). Our analyses thus demonstrate that although a globally euxinic deep ocean¹⁰ did not exist, Mesoproterozoic oceans were statistically more prone to euxinia than those of the Neoproterozoic.

We further estimated sedimentary sulfide generation through Earth's history. This property cannot be measured directly, but can be evaluated indirectly, because sulfide generated within sediments will bond with reactive iron to form pyrite. Hence, reactive iron acts as an effective sulfide sink, meaning that sulfide accumulation in pore waters and advective fluxes into marine waters—the free sulfide that would influence local animal ecology—will only occur in settings where most, if not all, highly reactive iron has been pyritized¹⁸. Thus, for shale deposited in oxic environments, pyrite contents broadly serve as a metric for total sulfide generation, and only environments with $\text{Fe}_{\text{pyr}}/\text{Fe}_{\text{hr}} > 0.70$ could have contained high levels of pore-water sulfide.

Analyses of the weight per cent iron in pyrite from oxic sediments (Fig. 1c) show an inverted pattern from Fig. 1b, with higher pyrite contents in the late Palaeoproterozoic/Mesoproterozoic bin, very low contents in the Neoproterozoic and Cambrian, and higher contents again in the Ordovician–Devonian (Kruskal–Wallis $\chi^2 = 25.44$, $P < 0.0001$; Supplementary Table 4). The Neoproterozoic captures a minimum in pyrite preservation that is about five times smaller than in modern oxic samples¹⁹. Similar results are seen for the proportion of oxic samples with inferred high levels of pore-water sulfide (Supplementary Table 4). It is worth emphasizing that the outlier is the Neoproterozoic—whether in the water column or the sediments, far more sulfide was generated in Mesoproterozoic and Palaeozoic basins.

These results have important implications for the physiology and oxygen tolerance of early animals, which probably began to diverge about 800 Myr ago²⁰. From observations in modern oxygen minimum

zones²¹ and experiments on sponges²², it has been suggested that early animals would have tolerated the low-oxygen conditions believed to characterize the Neoproterozoic era. With oxygen partially removed as a handbrake on earliest animal evolution, other inhibitors such as ambient sulfide²³ should be considered. Sulfide is a synergistic stressor in low-oxygen conditions because it binds to cytochrome oxidase and consequently inhibits aerobic respiration, lowering survival times under conditions of hypoxia²⁴. But in contrast to some modern oxygen minimum zones where sulfide often reaches the sediment–water

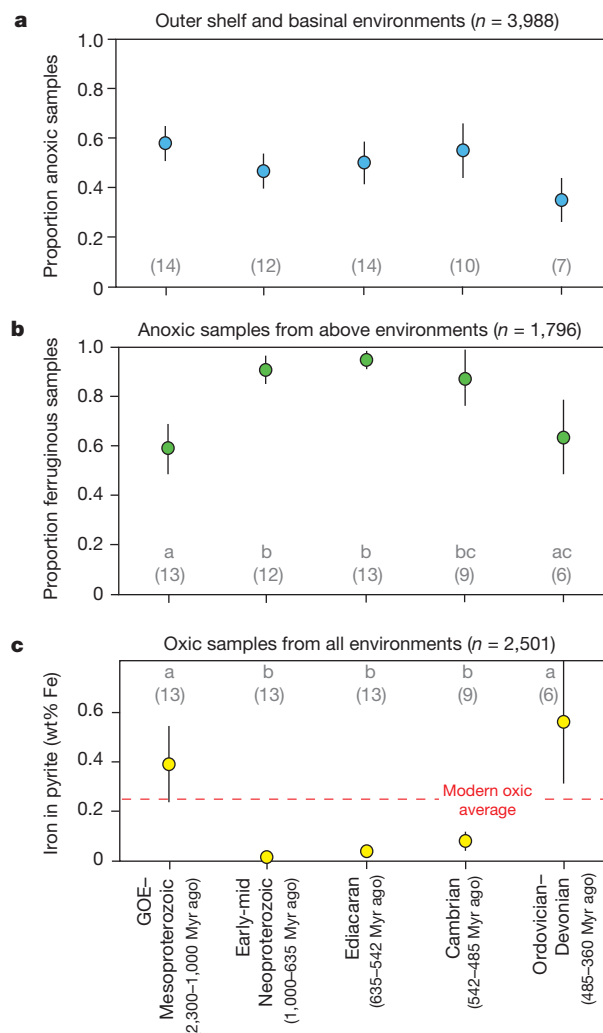


Figure 1 | Iron geochemical data compared using five time bins. The bins are: 2,300–1,000 Myr ago, 1,000–635 Myr ago, 635–542 Myr ago, 542–485 Myr ago and 485–360 Myr ago. The number of regions included in each bin is shown in grey text in parentheses. In **b** and **c**, the grey letters a, b or c represent the results of pairwise Wilcoxon tests. Bins joined by the same letter are not statistically significant ($P > 0.05$). **a**, The proportion of samples deposited beneath anoxic water columns ($\text{Fe}_{\text{hr}}/\text{Fe}_{\text{tot}} > 0.38$; refs 3, 19) from outer shelf and deep basin depositional environments. Each circle represents the average of regional proportions and the whiskers represent standard error. No bins are statistically different from one another (ANOVA $P = 0.54$; Kruskal–Wallis $P = 0.51$), and the proportion of oxic samples using a conservative threshold of $\text{Fe}_{\text{hr}}/\text{Fe}_{\text{tot}} < 0.22$ is also not significant (Supplementary Table 4). **b**, Proportion of samples deposited beneath ferruginous conditions from anoxic water columns ($\text{Fe}_{\text{hr}}/\text{Fe}_{\text{tot}} > 0.38$; $\text{Fe}_{\text{pyr}}/\text{Fe}_{\text{hr}} < 0.70$; ref. 3) from outer shelf and deep basin depositional environments. Each circle represents the average of regional proportions and whiskers represent standard error. **c**, Weight per cent iron in pyrite from samples deposited under oxic water columns from all depositional environments. Each circle represents the average of regional medians and whiskers represent standard error. The dashed line represents the modern oxic average from ref. 19. GOE, Great Oxidation Event.

interface, Neoproterozoic animals would have experienced little, if any, benthic sulfide flux. In fact, out of 1,243 oxic Neoproterozoic samples analysed, only 14 (about 1.1%) show possible evidence of pore-water sulfide. This bolsters suggestions that while earlier Neoproterozoic oceans may have prohibited large, metabolically active and carnivorous animals with higher oxygen demands, they could have accommodated early animals with small and thin body plans^{21,22}. Continued research on other proxies for the partial pressure of oxygen, p_{O_2} , will also help to place more precise constraints on early animal ecosystems²⁵.

These results raise the question of whether observed trends reflect biases in the data set, as there are known caveats when interpreting iron-speciation data, most prominently including the effects of weathering and diagenesis^{3,4,15} (Supplementary Information). However, as long as the data are sufficiently numerous, and geological and analytical biases are randomly distributed with respect to time, these processes will not affect our results (see ref. 26 regarding analogous errors in palaeobiological data). The impact of random and systematic error can be tested with resampling and sensitivity analyses. Sensitivity analyses excluding possibly inappropriate samples and regions of low data coverage, and a further analysis using only Mesoproterozoic (1,600–1,000 Myr ago) samples for the oldest time bin, are consistent with results from the entire data set (Supplementary Table 4). Further, in synthetically re-sampled data sets, the Cambrian distribution of anoxic samples is indistinguishable from the Ediacaran distribution (Supplementary Fig. 2). To test whether inappropriate binning may contribute to the invariance in Fig. 1a, data from each region from the interval of 800–360 Myr ago were plotted individually with respect to time (Fig. 2). Although there is clear spatial heterogeneity (as in the modern ocean), there are no apparent ‘oxygenation events’, and a linear regression is not significant ($P = 0.45$; see also a local regression (LOESS) of geographically unbinned data, Supplementary Fig. 1).

It has been argued that trace metals in anoxic shales capture the spatial contraction of basinal anoxia across the Ediacaran–Cambrian transition^{1,5,6}, probably driven by increasing atmospheric p_{O_2} . To evaluate the consistency between iron-speciation and trace-metal results, we re-analysed a well vetted sedimentary uranium data set⁶ using statistical methods similar to those employed in the iron analyses, although lower data density precludes a basin-normalized approach. The maximum ratios of metals to total organic carbon (TOC) are often taken as a guide to the metal inventory in ancient seawater; however, without a priori knowledge of basin restriction and secondary mineralization or local redistribution for each sample^{1,27}, statistical approaches based on the entire population of data are appropriate.

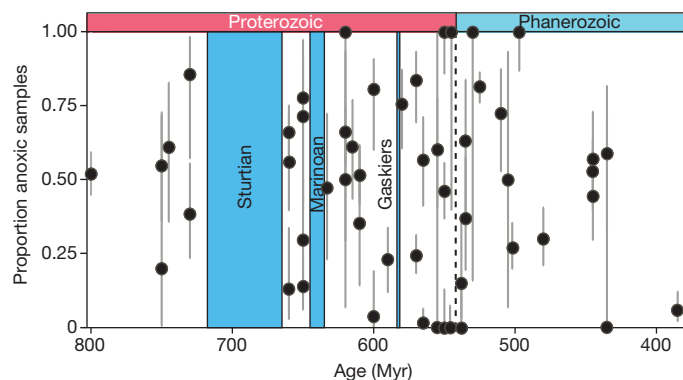


Figure 2 | Unbinned analysis of the proportion of anoxic samples from each region for the time period 800–360 Myr ago. Ages for different regions based on best geological estimates; Neoproterozoic samples from the same region were separated based on the global Sturtian and Marinoan glaciations, the Gaskiers glaciation or the mid-Ediacaran Shuram carbon isotope excursion and its equivalents, and the Ediacaran–Cambrian boundary. Grey bars represent 95% binomial confidence intervals.

When anoxic, organic-rich shales (TOC > 0.4%) are binned into Neoproterozoic, Cambrian–Silurian and Devonian–Permian domains, uranium/TOC significantly increases with younger age (Kruskal–Wallis $\chi^2 = 75.53$, $P < 0.0001$; all pairwise Wilcoxon tests $P < 0.0001$; see Supplementary Table 5). The Devonian–Permian time bin contains a much higher number of enriched outlier values relative to the Cambrian–Silurian time bin (Supplementary Fig. 3). Thus, while the uranium/TOC record does show a punctuated increase in oxygenation at the Ediacaran–Cambrian boundary, it is also consistent with iron geochemical data (Fig. 1a) that suggest that full oxygenation of the oceans did not occur until later.

The question then becomes the magnitude of oxygenation implied by the iron and trace-metal data sets. Recent models indicate that relatively subtle changes in seafloor anoxia and the proportion of the sea floor that was ferruginous rather than euxinic will lead to dramatic changes in seawater trace-metal inventories, and by inference, trace-metal enrichments in shales^{1,5,6,28}. Trace-metal enrichments thus respond to the total size of anoxic sinks, whereas the binned iron data are tracking the percentage of sediments sampled in the stratigraphic record bathed by anoxic waters. As large changes in anoxic sink size can manifest as small shifts in the percentage of anoxic sea floor, we propose that trace-metal abundances and the binned iron-speciation records are complementary but have different thresholds; that is, binned iron data require a larger change in global oxygen to record a statistically significant (see above) signal.

Although absolute values of p_{O_2} in the geological record are notoriously difficult to track, the iron-speciation database results constrain the magnitude of the latest Proterozoic p_{O_2} increase indicated by trace-metal compilations. Canfield¹⁰ earlier posited that at atmospheric $p_{O_2} < 30$ –40% PAL (Present Atmospheric Level), deeper water masses tend towards anoxia, albeit dependent upon phosphorus fluxes. Although this was intended to constrain oxygen levels before Ediacaran oxygenation, it also provides an upper bound on Cambrian p_{O_2} , given the lack of statistical change through time. The distribution of animals in modern oceans^{8,29} suggests that the Cambrian metazoans recorded by fossils required oxygen levels above about 10% PAL, but not much more than that, given that equally large, mobile and skeletonized animals live at and even below this level in the modern ocean^{8,29}. The combined constraints from iron-speciation and palaeontological data are therefore consistent with molybdenum isotope data²⁸, global sedimentary sulfate reduction rates³⁰, uranium/TOC⁶ (Supplementary Table 5) and some models of atmospheric oxygen through time³¹. All offer evidence that oxygenation of the ocean–atmosphere system to essentially modern levels and a persistently oxygenated deep ocean is in large part a post-Cambrian phenomenon, as has been separately hypothesized for black shale distribution³². Overall, these analyses imply a modest increase in oxygen during the Ediacaran and Cambrian (Fig. 3).

This evolving picture of Earth’s redox state would seem to diminish the impact of oxygen as a causal factor in Cambrian animal radiation. Observations from modern oxygen minimum zones, however, suggest that a small increase in p_{O_2} could still be a critical environmental trigger owing to nonlinear threshold effects at very low oxygen levels. Many important ecological responses for macrofaunal organisms, including feeding efficiency⁹, species-level diversity⁸, and carnivore abundance and species richness⁷ exhibit threshold changes in the range of 5–20 μM oxygen, or ~ 2 –7% of modern surface ocean oxygen concentrations—results that are strikingly similar to the changes accommodated by this analysis. Thus, a relatively small increase in p_{O_2} could reasonably have moved animals past critical ecological thresholds, especially with respect to carnivory⁷, which might have driven Cambrian diversification. It remains possible, though, that sufficient oxygen for large, muscular carnivores existed before the Cambrian (Fig. 3). The critical question is whether oxygen availability before the Ediacaran–Cambrian transition was in the ~ 1 –5% PAL range (at which modern animal ecology is severely limited), or higher.

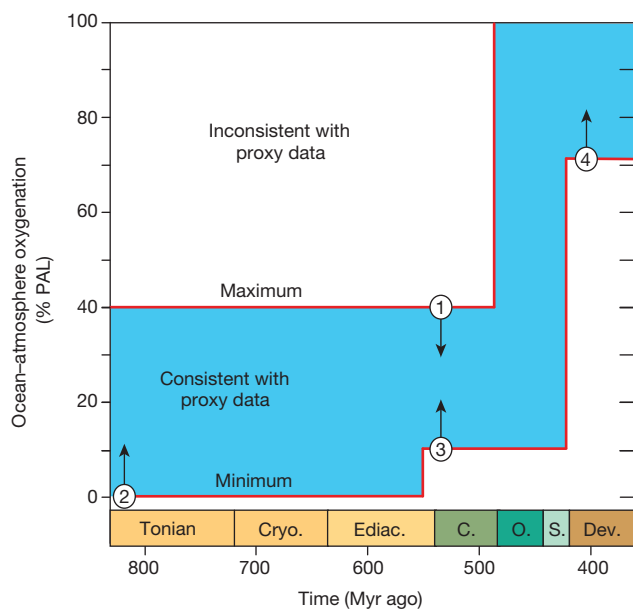


Figure 3 | Ocean–atmosphere oxygenation through the Proterozoic–Phanerozoic transition. Data are based on the combined absence of a statistically significant oxygenation event in iron-speciation data and the presence of an oxygenation event in redox-sensitive trace-metal inventories. Oxygen constraints include: (1) persistently anoxic subsurface waters requiring less than 40% PAL¹⁰ (iron-speciation data for the Ordovician–Devonian is not statistically different from that of previous time bins, but data are sparse and may be subject to sampling biases—see Supplementary Information); (2) a minimum oxygen level of ~0.5–1% PAL, required for the appearance of mass-dependent sulfur isotope fractionation, red beds, and the earliest animals^{1,21}, although oxygen levels before ~810 Myr ago may have been lower²⁵; (3) oxygen levels exceeding 10% PAL^{8,29}, required by the Cambrian biota; and (4) oxygen levels must have exceeded 70% PAL in the latest Silurian, as deduced from the presence of fires³³. Within these constraints, oxygenation could have followed many different paths, but full oxygenation of the ocean–atmosphere system is a Palaeozoic phenomenon. Ediac., Ediacaran; Cryo., Cryogenian; C., Cambrian; O., Ordovician; S., Silurian; Dev., Devonian.

Coupled with other geochemical data, our global database of iron-speciation measurements provides an increasingly resolved and quantitative picture of redox evolution in Proterozoic and Palaeozoic oceans. These data point to proportionally higher basinal euxinia in Mesoproterozoic and younger Palaeozoic basins, with sediment and water-column sulfide generation reaching a minimum in the Neoproterozoic oceans. Ediacaran oxygenation was relatively modest, but may have been sufficient to remove environmental barriers to Cambrian animal evolution. Future sedimentary geochemical sampling of both iron and redox-sensitive trace-metal data will increase temporal resolution and the power of inference tests, with statistical analysis in a basin-normalized context providing more robust hypotheses of deep-time global change.

Received 26 September 2014; accepted 21 May 2015.

- Lyons, T. W., Reinhard, C. T. & Planavsky, N. J. The rise of oxygen in Earth's early ocean and atmosphere. *Nature* **506**, 307–315 (2014).
- Lenton, T. M., Boyle, R. A., Poulton, S. W., Shields-Zhou, G. A. & Butterfield, N. J. Co-evolution of eukaryotes and ocean oxygenation in the Neoproterozoic era. *Nature Geosci.* **7**, 257–265 (2014).
- Poulton, S. W. & Canfield, D. E. Ferruginous conditions: a dominant feature of the ocean through Earth's history. *Elements* **7**, 107–112 (2011).
- Lyons, T. W. & Severmann, S. A critical look at iron paleoredox proxies: new insights from modern euxinic marine basins. *Geochim. Cosmochim. Acta* **70**, 5698–5722 (2006).
- Reinhard, C. T. *et al.* Proterozoic ocean redox and biogeochemical stasis. *Proc. Natl Acad. Sci. USA* **110**, 5357–5362 (2013).

- Partin, C. A. *et al.* Large-scale fluctuations in Precambrian atmospheric and oceanic oxygen levels from the record of U in shales. *Earth Planet. Sci. Lett.* **369–370**, 284–293 (2013).
- Sperling, E. A. *et al.* Oxygen, ecology, and the Cambrian radiation of animals. *Proc. Natl Acad. Sci. USA* **110**, 13446–13451 (2013).
- Levin, L. A. Oxygen Minimum Zone benthos: adaptation and community response to hypoxia. *Oceanogr. Mar. Biol. Annu. Rev.* **41**, 1–45 (2003).
- Wouds, C. *et al.* Oxygen as a control on seafloor biological communities and their roles in sedimentary carbon cycling. *Limnol. Oceanogr.* **52**, 1698–1709 (2007).
- Canfield, D. E. A new model for Proterozoic ocean chemistry. *Nature* **396**, 450–453 (1998).
- Shen, Y., Knoll, A. H. & Walter, M. R. Evidence for low sulphate and anoxia in a mid-Proterozoic marine basin. *Nature* **423**, 632–635 (2003).
- Canfield, D. E., Poulton, S. W. & Narbonne, G. M. Late-Neoproterozoic deep-ocean oxygenation and the rise of animal life. *Science* **315**, 92–95 (2007).
- Planavsky, N. J. *et al.* Widespread iron-rich conditions in the mid-Proterozoic ocean. *Nature* **477**, 448–451 (2011).
- Sperling, E. A. *et al.* Redox heterogeneity of subsurface waters in the Mesoproterozoic ocean. *Geobiology* **12**, 373–386 (2014).
- Canfield, D. E. *et al.* Ferruginous conditions dominated later Neoproterozoic deep-water chemistry. *Science* **321**, 949–952 (2008).
- Johnston, D. T. *et al.* Searching for an oxygenation event in the fossiliferous Ediacaran of northwestern Canada. *Chem. Geol.* **362**, 273–286 (2013).
- Alroy, J. *et al.* Phanerozoic trends in the global diversity of marine invertebrates. *Science* **321**, 97–100 (2008).
- Canfield, D. E. Reactive iron in marine sediments. *Geochim. Cosmochim. Acta* **53**, 619–632 (1989).
- Raiswell, R. & Canfield, D. E. Sources of iron for pyrite formation in marine sediments. *Am. J. Sci.* **298**, 219–245 (1998).
- Erwin, D. H. *et al.* The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science* **334**, 1091–1097 (2011).
- Sperling, E. A., Halverson, G. P., Knoll, A. H., Macdonald, F. A. & Johnston, D. T. A basin redox transect at the dawn of animal life. *Earth Planet. Sci. Lett.* **371–372**, 143–155 (2013).
- Mills, D. B. *et al.* Oxygen requirements of the earliest animals. *Proc. Natl Acad. Sci. USA* **111**, 4168–4172 (2014).
- Martin, W. *et al.* Early cell evolution, eukaryotes, anoxia, sulfide, oxygen, fungi first (?), and a tree of genomes revisited. *IUBMB Life* **55**, 193–204 (2003).
- Vaquier-Sunyer, R. & Duarte, C. M. Sulfide exposure accelerates hypoxia-driven mortality. *Limnol. Oceanogr.* **55**, 1075–1082 (2010).
- Planavsky, N. J. *et al.* Low Mid-Proterozoic atmospheric oxygen levels and the delayed rise of animals. *Science* **346**, 635–638 (2014).
- Sepkoski, J. J. Jr. Ten years in the library: new data confirm paleontological patterns. *Paleobiology* **19**, 43–51 (1993).
- Algeo, T. J. & Rowe, H. Paleocyanographic applications of trace-metal concentration data. *Chem. Geol.* **324–325**, 6–18 (2011).
- Dahl, T. W. *et al.* Devonian rise in atmospheric oxygen correlated to the radiation of terrestrial plants and large predatory fish. *Proc. Natl Acad. Sci. USA* **107**, 17911–17915 (2010).
- Rhoads, D. C. & Morse, J. W. Evolutionary and ecologic significance of oxygen-deficient marine basins. *Lethaia* **4**, 413–428 (1971).
- Leavitt, W. D., Halevy, I., Bradley, A. S. & Johnston, D. T. Influence of sulfate reduction rates on the Phanerozoic sulfur isotope record. *Proc. Natl Acad. Sci. USA* **110**, 11244–11249 (2013).
- Bergman, N. M., Lenton, T. M. & Watson, A. J. COPSE: a new model of biogeochemical cycling over Phanerozoic time. *Am. J. Sci.* **304**, 397–437 (2004).
- Berry, W. B. & Wilde, P. Progressive ventilation of the oceans; an explanation for the distribution of the lower Palaeozoic black shales. *Am. J. Sci.* **278**, 257–275 (1978).
- Belcher, C. M. & McElwain, J. C. Limits for combustion in low O₂ redefine paleoatmospheric predictions for the Mesozoic. *Science* **321**, 1197–1200 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Cole, A. Masterson, E. Beirne, G. Resendiz, A. Miller and W. Mai for laboratory assistance, E. Smith, T. Petach, M. Laflamme, S. Darroch, P. Myrow, J. Strauss, C. Carbone and G. Narbonne for field assistance, L. Levin, P. Girguis, C. Frieder, L. Duncan, C. Partin, S. Darroch and U. Farrell for discussions, and E. Hammarlund, C. Scott, R. Raiswell, G. Gilleaudeau, S. Sahoo, Y. Shen, J. Creveling, T. Dahl, C. Partin and D. Stolper for providing raw data from published studies. E.A.S. is funded by Agouron Geobiology and NAI Postdoctoral Fellowships and by the NSF-EAR 1324095 grant to L. Levin. This work is supported by the NSF-EAR 1324095 grant to D.T.J. and by the NASA Astrobiology Institute (A.H.K. and F.A.M.).

Author Contributions E.A.S. and D.T.J. conceived the project, E.A.S., A.S.M., B.C.G., G.P.H., F.A.M. and D.T.J. collected samples, E.A.S., A.S.M., B.C.G. and M.K. completed geochemical measurements, E.A.S. and C.J.W. compiled data and analysed the global data set, and E.A.S. wrote the paper with input from all co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.A.S. (esper@stanford.edu) or D.T.J. (johnston@eps.harvard.edu).

The ancestry and affiliations of Kennewick Man

Morten Rasmussen^{1,2}, Martin Sikora^{1*}, Anders Albrechtsen^{3*}, Thorfinn Sand Korneliussen^{1*}, J. Víctor Moreno-Mayar^{1*}, G. David Poznik⁴, Christoph P. E. Zollikofer⁵, Marcia S. Ponce de León⁵, Morten E. Allentoft¹, Ida Moltke³, Hákon Jónsson¹, Cristina Valdiosera^{1,6}, Ripan S. Malhi⁷, Ludovic Orlando¹, Carlos D. Bustamante^{2,8}, Thomas W. Stafford Jr^{1,9}, David J. Meltzer¹⁰, Rasmus Nielsen^{1,11} & Eske Willerslev¹

Kennewick Man, referred to as the Ancient One by Native Americans, is a male human skeleton discovered in Washington state (USA) in 1996 and initially radiocarbon dated to 8,340–9,200 calibrated years before present (BP)¹. His population affinities have been the subject of scientific debate and legal controversy. Based on an initial study of cranial morphology it was asserted that Kennewick Man was neither Native American nor closely related to the claimant Plateau tribes of the Pacific Northwest, who claimed ancestral relationship and requested repatriation under the Native American Graves Protection and Repatriation Act (NAGPRA). The morphological analysis was important to judicial decisions that Kennewick Man was not Native American and that therefore NAGPRA did not apply. Instead of repatriation, additional studies of the remains were permitted². Subsequent craniometric analysis affirmed Kennewick Man to be more closely related to circumpacific groups such as the Ainu and Polynesians than he is to modern Native Americans². In order to resolve Kennewick Man's ancestry and affiliations, we have sequenced his genome to $\sim 1\times$ coverage and compared it to worldwide genomic data including for the Ainu and Polynesians. We find that Kennewick Man is closer to modern Native Americans than to any other population worldwide. Among the Native American groups for whom genome-wide data are available for comparison, several seem to be descended from a population closely related to that of Kennewick Man, including the Confederated Tribes of the Colville Reservation (Colville), one of the five tribes claiming Kennewick Man. We revisit the cranial analyses and find that, as opposed to genome-wide comparisons, it is not possible on that basis to affiliate Kennewick Man to specific contemporary groups. We therefore conclude based on genetic comparisons that Kennewick Man shows continuity with Native North Americans over at least the last eight millennia.

The skeleton of Kennewick Man was inadvertently discovered in July of 1996 in shallow water along the Columbia River shoreline outside Kennewick, Washington state, USA. On several visits to the locality over the following month, some 300 bone elements and fragments were collected, ultimately comprising $\sim 90\%$ of an adult male human skeleton³. The initial assessment of this individual was that he was a historic-period Euro-American, based largely on his apparently “Caucasoid-like”³ cranium, along with a few artefacts found nearby (later proved not to be associated with the skeletal remains). However, radiocarbon dating subsequently put the age of the skeleton in the Early Holocene⁴. The claim that Kennewick Man was anatomically distinct from modern Native Americans in general, and in particular from those tribes inhabiting northwest North America⁴, sparked a legal battle over the disposition of the skeletal remains. Five tribes who inhabit that

region requested the remains be returned to them for reburial under the Native American Graves Protection and Repatriation Act (NAGPRA). The US Army Corps of Engineers, which manages the land where Kennewick Man was found, announced their intent to do so. That in turn prompted a lawsuit to block the repatriation^{2,5}, and generated considerable scientific controversy as to Kennewick Man's ancestry and affinities (for example, refs 3, 6–9). The lawsuit ultimately (in 2004) resulted in a judicial ruling in favour of a detailed study of the skeletal remains, the results of which were recently published².

These studies provide important details on, for example, Kennewick Man's life history, refine his antiquity to $8,358 \pm 21$ ¹⁴C years BP or to within a two sigma range of 8,400–8,690 calibrated years BP (based on 90% marine diet, and 750 year marine reservoir correction), and demonstrate that the body had been intentionally buried and had eroded out shortly before discovery². They also include anatomical and morphometric analyses, which confirm earlier studies that Kennewick Man resembles circumpacific populations, particularly the Ainu and Polynesians^{2,10}; that he has certain “European-like morphological” traits²; and that he is anatomically distinct from modern Native Americans². These results are interpreted as indicating that Kennewick Man was a descendant of a population that migrated earlier than, and independently of, the population(s) that gave rise to modern Native Americans².

However, those recent studies did not include DNA analysis. Herein we present the genome sequence of Kennewick Man in order to resolve his ancestry and affinities with modern Native Americans. There were several prior efforts to recover genetic material from Kennewick Man¹¹, but none were successful.

We obtained $\sim 1\times$ coverage of the genome, from 200 mg of metacarpal bone specimen (Supplementary Information 1) using previously published methods^{12,13}. The endogenous DNA content was between 0.4% and 1.4% for double-stranded and single-stranded libraries, respectively (Supplementary Information 2). Average fragment length was 53.6 base pairs (bp) and the sample exhibited damage patterns typical of ancient DNA, with excessive deamination of cytosine towards the ends of the fragments (Supplementary Information 2). Similarly, patterns of DNA decay agree with published expectations¹⁴, and display an estimated molecular half-life corresponding to 3,670 years for 100-bp molecules (Supplementary Information 3). The mitochondrial genome was sequenced to $\sim 71\times$ coverage and is placed at the root of haplogroup X2a (Extended Data Fig. 1 and Supplementary Information 2), and the Y-chromosome haplogroup is Q-M3 (Extended Data Fig. 2 and Supplementary Information 5); both uniparental lineages are found almost exclusively among contemporary Native Americans^{15,16}. We used the X chromosome to

¹Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, DK-1350 Copenhagen K, Denmark. ²Department of Genetics, School of Medicine, Stanford University, Littlefield Center, Stanford, California 94305, USA. ³The Bioinformatics Centre, Department of Biology, University of Copenhagen, Ole Maaloes Vej 5, DK-2200 Copenhagen N, Denmark.

⁴Program in Biomedical Informatics, Stanford University, Stanford, California 94305, USA. ⁵Anthropological Institute, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland.

⁶Department of Archaeology and History, La Trobe University, Melbourne, Victoria 3086, Australia. ⁷Department of Anthropology and Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, 209F Davenport Hall, 607 Matthews Avenue, Urbana, Illinois 61801, USA. ⁸Center for Evolutionary and Human Genomics, Stanford University, Littlefield Center, Stanford, California 94305, USA. ⁹AMS, ¹⁴C Dating Centre, Department of Physics & Astronomy, University of Aarhus, Ny Munkegade 120, DK-8000 Aarhus C, Denmark. ¹⁰Department of Anthropology, Southern Methodist University, Dallas, Texas 75275, USA. ¹¹Department of Integrative Biology, University of California, Berkeley, 4134 Valley Life Sciences Building, Berkeley, California 94720, USA.

*These authors contributed equally to this work.

conservatively estimate contamination to be 2.5%, which is within the normal range obtained observed in genomic data from ancient human remains¹⁷, and we further show this contamination to be of European origin (Supplementary Information 4).

We compiled an autosomal reference data set consisting of published SNP array data^{18–23} as well as new data generated from one of the claimant tribes, the Colville (Supplementary Information 10). Due to high levels of recent admixture in many Native American populations, we masked European ancestry from the Native Americans (Supplementary Information 6). No masking was done on the Kennewick Man. When we compare Kennewick Man with the worldwide panel of populations, a clear genetic similarity to Native Americans is observed both in principal components analysis (PCA) and using f_3 -outgroup statistics (Fig. 1a, b). In particular, we can reject the hypothesis that Kennewick Man is more closely related to Ainu or Polynesians than he is to Native Americans, as seen in a D -statistic-based test where no trees of the type ((CHB,Ainu/Polynesian),(X,Karitiana)) with X being Kennewick Man, the Clovis age Anzick-1 child (ref. 12) or a modern Native American genome are rejected (Extended Data Fig. 3). Model-based clustering using ADMIXTURE²⁴ shows that Kennewick Man has ancestry proportions most similar to those of other Northern Native Americans (Fig. 1c and Supplementary Information 7), especially the Colville, Ojibwa, and Algonquin. Considering the Americas only, f_3 -outgroup and D -statistic based analyses show that Kennewick Man, like the Anzick-1 child, shares a high degree of ancestry with Native Americans from Central and South America, and that Kennewick Man also groups with geographically close tribes including the Colville (Fig. 2a, b and Extended Data Fig. 4). Despite this similarity, Anzick-1 and Kennewick Man have dissimilar genetic affinities to contemporary Native Americans. In particular, we find that Anzick-1 is more closely related to Central/Southern Native Americans than is Kennewick Man (Extended Data Fig. 5). The pattern observed in Kennewick Man is mirrored in the Colville, who also show a high affinity with Southern populations (Fig. 2c), but are most closely related to a neighbouring population in the data set (Stswecem'c; Extended

Data Fig. 4c). This is in contrast to other populations such as the Chipewyan, who are more closely related to Northern Native Americans rather than to Central/Southern Native Americans in all comparisons (Fig. 2d and Extended Data Fig. 4d).

Our results are in agreement with a basal divergence of Northern and Central/Southern Native American lineages as suggested from the analysis of the Anzick-1 genome¹². However, the genetic affinities of Kennewick Man reveal additional complexity in the population history of the Northern lineage. The finding that Kennewick is more closely related to Southern than many Northern Native Americans (Extended Data Fig. 4) suggests the presence of an additional Northern lineage that diverged from the common ancestral population of Anzick-1 and Southern Native Americans (Fig. 3). This branch would include both Colville and other tribes of the Pacific Northwest such as the Stswecem'c, who also appear symmetric to Kennewick with Southern Native Americans (Extended Data Fig. 4). We also find evidence for additional gene flow into the Pacific Northwest related to Asian populations (Extended Data Fig. 5), which is likely to post-date Kennewick Man. We note that this gene flow could originate from within the Americas, for example in association with the migration of paleo-Eskimos or Inuit ancestors within the past 5,000 years²⁵, or the gene flow could be post-colonial¹⁹.

We used a likelihood ratio test to test for direct ancestry of Kennewick Man for two members of the Colville tribe who show no evidence of recent European admixture. This test allows us to determine if the patterns of allele frequencies in the Colville and Kennewick Man are compatible with direct ancestry of the Colville from the population to which Kennewick Man belonged, without any additional gene flow. As a comparison we also included analyses of four other Native Americans with high quality genomes: two Northern Athabaskan individuals from Canada²⁵ and two Karitiana individuals from Brazil^{12,13}. Although the test rejects the null hypothesis of direct ancestry with no subsequent gene flow in all cases, it only does so very weakly for the Colville tribe members (Table 1 and Supplementary Information 8). These findings can be explained as: (1) the Colville

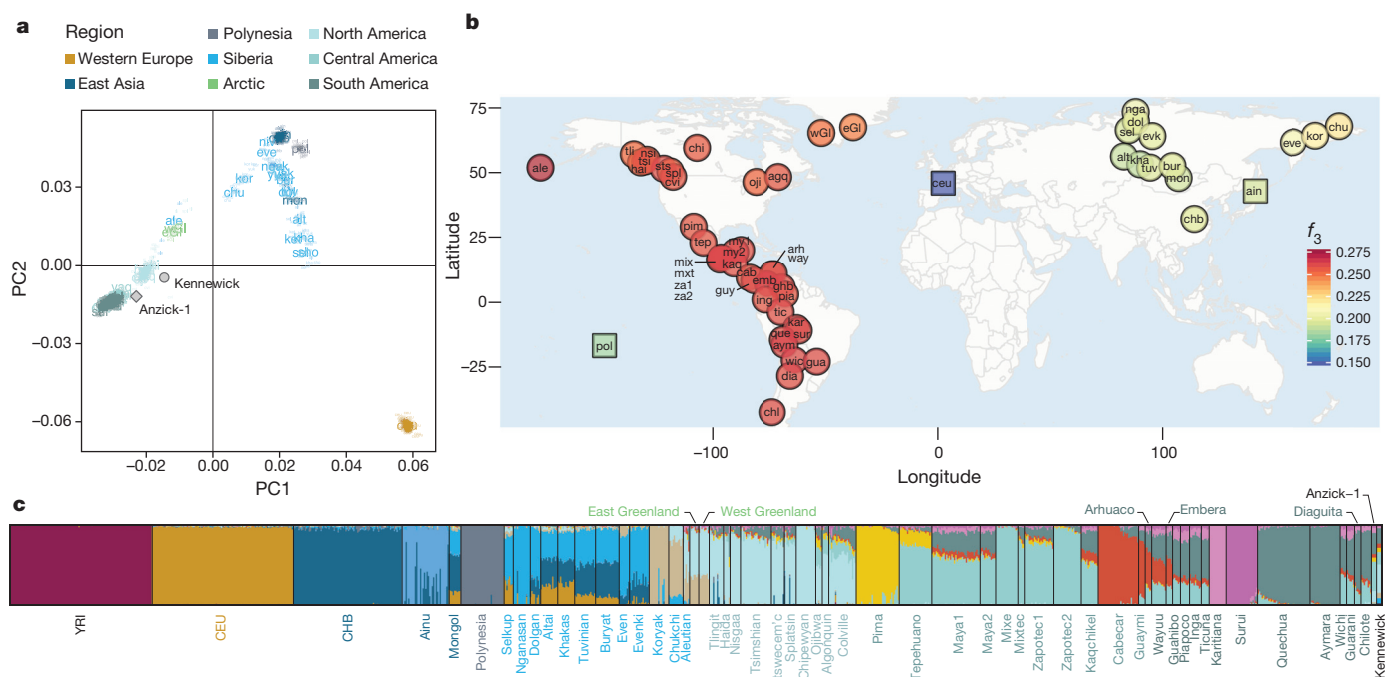


Figure 1 | Genetic affinities between Kennewick Man and a panel of world-wide populations. **a**, Principal components analysis (PCA) projecting Kennewick Man and Anzick-1 onto a set of out-of-Africa populations. **b**, Heat map of f_3 -outgroup statistics between Kennewick Man, Native Americans, Siberians and additional populations with suggested relationship to Kennewick

Man (in squares). Warmer colours indicate higher allele sharing. For list of population labels, see the Methods section. **c**, Admixture proportions for worldwide set of population, including masked Native American, Anzick-1 and Kennewick, shown at $K = 14$.

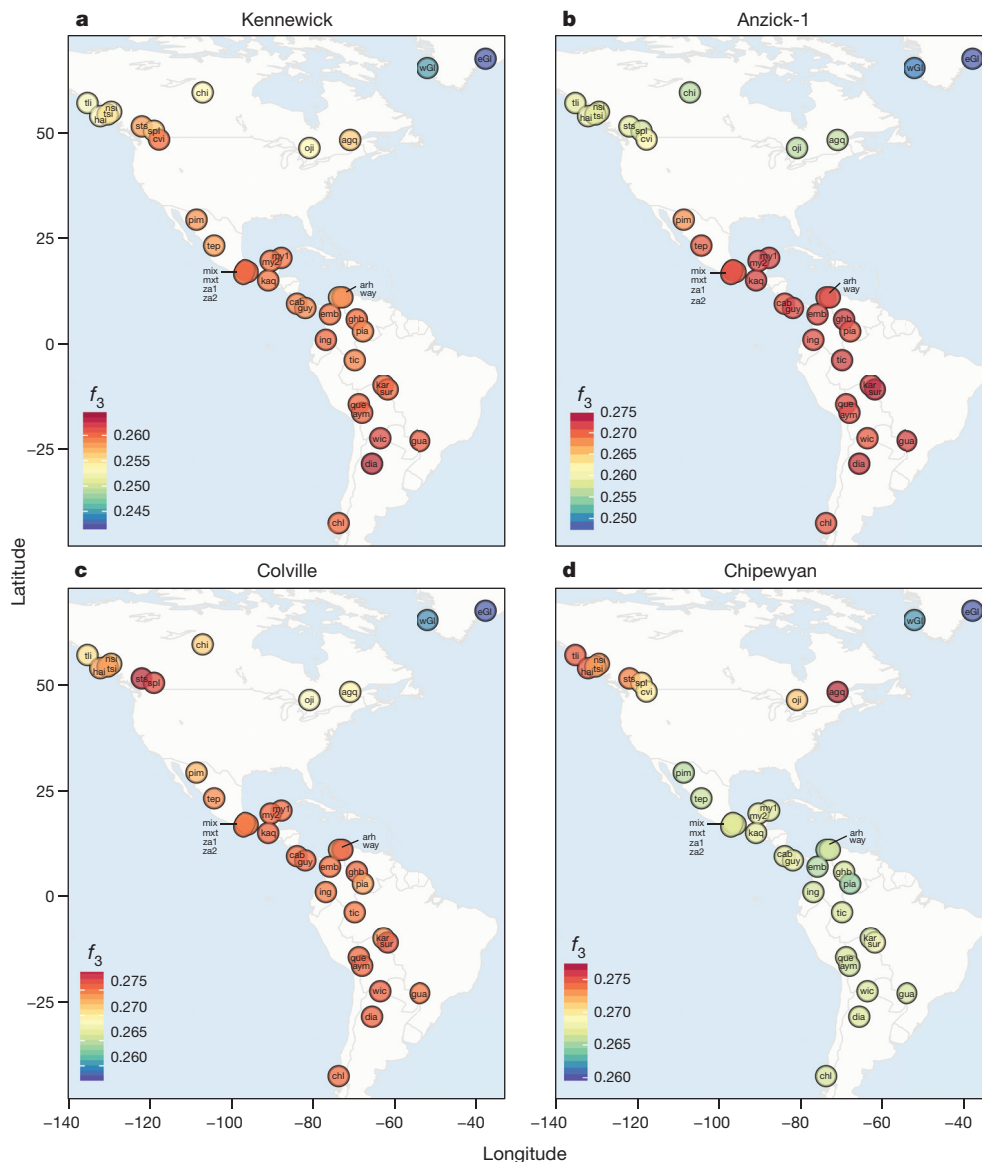


Figure 2 | Shared ancestry among samples within the Americas. a–d, Heat maps of f_3 -outgroup statistics testing (YRI; Native Americans, X), where X is Kennewick Man (a), Anzick-1 (b), Colville (c) or Chipewyan (d). Warmer colours indicate higher allele sharing, for list of population labels, see the Methods section.

individuals are direct descendants of the population to which Kennewick Man belonged, but subsequently received some relatively minor gene flow from other American populations within the last ~8,500 years, in agreement with our findings above; (2) the Colville individuals descend from a population that ~8,500 years was slightly

diverged from the population which Kennewick Man belonged or (3) a combination of both.

It has been asserted that “...cranial morphology provides as much insight into population structure and affinity as genetic data”²². However, although recent and previous craniometric analyses have consistently concluded that Kennewick Man is unlike modern Native Americans, they disagree regarding his closest population affinities, the cause of the apparent differences between Kennewick Man and modern Native Americans, and whether the differences are historically important (for example, represent an earlier, separate

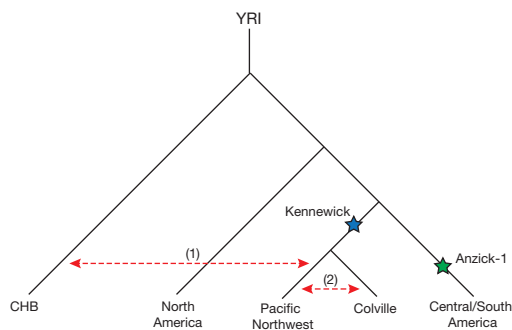


Figure 3 | Illustration of Native American population history. Depicted is a population tree consistent with the broad affinities between modern and ancient Native Americans. Kennewick Man and the Anzick-1 child are indicated with blue and green stars respectively. Red dashed arrows indicate gene flow (1) of Asian-related ancestry with tribes of the Pacific Northwest and (2) between Colville and neighbouring tribes.

Table 1 | Direct ancestry test

	Coalescence probability in Kennewick lineage (c_1)	Coalescence probability in reference lineage (c_2)	$2 \times \text{Log likelihood ratio of } H_0: c_1 = 0 \text{ vs } H_A: c_1 > 0$
Colville 2	0.015	0.072	19.41
Colville 8	0.007	0.097	3.93
Athabascan 1	0.048	0.073	505.52
Athabascan 2	0.056	0.097	807.69
BI16 (Karitiana)	0.040	0.140	423.87
HGDP00998 (Karitiana)	0.040	0.170	446.30

c_1 is the probability of coalescence in the Kennewick lineage and c_2 is the probability of coalescence in the reference population lineage. A value of $c_1 = 0$ corresponds to direct Kennewick ancestry of the reference population with no subsequent gene flow. Smaller likelihood ratios indicate less evidence against direct Kennewick ancestry.

migration to the Americas), or simply represent intra-population variation^{2,3,7,10,26–28}. These inconsistencies are probably owing to the difficulties in assigning a single individual when comparing to population-mean data, without explicitly taking into account within-population variation. Reanalysis of W. W. Howells' worldwide modern human craniometric data set²⁹ (Supplementary Information 9) shows that biological population affinities of individual specimens cannot be resolved with any statistical certainty. Although our individual-based craniometric analyses confirm that Kennewick Man tends to be more similar to Polynesian and Ainu peoples than to Native Americans, Kennewick Man's pattern of craniometric affinity falls well within the range of affinity patterns evaluated for individual Native Americans (Supplementary Information 9). For example, the Arikara from North Dakota (the Native American tribe representing the geographically closest population in Howells' data set to Kennewick), exhibit with high frequency closest affinities with Polynesians (Supplementary Information 9). Yet, the Arikara have typical Native-American mitochondrial DNA haplogroups³⁰, as does Kennewick Man. We conclude that the currently available number of independent phenetic markers is too small, and within-population craniometric variation too large, to permit reliable reconstruction of the biological population affinities of Kennewick Man.

In contrast, block bootstrap results from the autosomal DNA data are highly statistically significant (Extended Data Fig. 3), showing stronger association of the Kennewick man with Native Americans than with any other continental group. We also observe that the autosomal DNA, mitochondrial DNA and Y chromosome data all consistently show that Kennewick Man is directly related to contemporary Native Americans, and thus show genetic continuity within the Americas over at least the past 8,000 years. Identifying which modern Native American groups are most closely related to Kennewick Man is not possible at this time as our comparative DNA database of modern peoples is limited, particularly for Native-American groups in the United States. However, among the groups for which we have sufficient genomic data, we find that the Colville, one of the Native American groups claiming Kennewick Man as ancestral, show close affinities to that individual or at least to the population to which he belonged. Additional modern descendants could be identified as more Native American groups are sequenced. Finally, it is clear that Kennewick Man differs significantly from the Anzick-1 child who is more closely related to the modern tribes of Mesoamerica and South America¹², possibly suggesting an early population structure within the Americas.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 April; accepted 10 June 2015.

Published online 18 June 2015.

1. Taylor, R. E., Kirner, D. L., Southon, J. R. & Chatters, J. C. Radiocarbon dates of Kennewick Man. *Science* **280**, 1171 (1998).
2. Owsley, D. W. & Jantz, R. L. *Kennewick Man* (Texas A&M Univ. Press, 2014).
3. Chatters, J. C. The recovery and first analysis of an Early Holocene human skeleton from Kennewick, Washington. *Am. Antiq.* **65**, 291–316 (2000).
4. Chatters, J. C. *Ancient Encounters* (Simon and Schuster, 2001).
5. Bonnichen, V. *United States*, 217 F. Supp. 2d 1116 (D. Or. 2002).
6. Owsley, D. W. & Jantz, R. L. Archaeological politics and public interest in paleoamerican studies: lessons from Gordon Creek Woman and Kennewick Man. *Am. Antiq.* **66**, 565–575 (2001).
7. Powell, J. F. *The First Americans* (Cambridge Univ. Press, 2005).
8. Swedlund, A. & Anderson, D. Gordon Creek Woman meets Kennewick Man: new interpretations and protocols regarding the peopling of the Americas. *Am. Antiq.* **64**, 569–576 (1999).
9. Thomas, D. H. *Skull Wars: Kennewick Man, Archaeology, and the Battle for Native American Identity* (Basic Books, 2001).
10. Steele, D. G. & Powell, J. F. in *The First Americans* (ed. Jablonski, N. G.) (Univ. California Press, 2002).
11. McManamon, F. Kennewick Man. <http://www.nps.gov/archeology/kennewick/Index.htm> (National Park Service, Washington DC, USA, 2004).
12. Rasmussen, M. et al. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225–229 (2014).


13. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
14. Allentoft, M. E. et al. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. R. Soc. Lond. B* **279**, 4724–4733 (2012).
15. Perego, U. A. et al. The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. *Genome Res.* **20**, 1174–1179 (2010).
16. Dulik, M. C. et al. Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athapaskan- and Eskimoan-speaking populations. *Proc. Natl Acad. Sci. USA* **109**, 8471–8476 (2012).
17. Seguin-Orlando, A. et al. Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113–1118 (2014).
18. Reich, D. et al. Reconstructing Native American population history. *Nature* **488**, 370–374 (2012).
19. Verdu, P. et al. Patterns of admixture and population structure in native populations of northwest North America. *PLoS Genet.* **10**, e1004530 (2014).
20. Frazer, K. A. et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
21. Japanese Archipelago Human Population Genetics Consortium. The history of human populations in the Japanese Archipelago inferred from genome-wide SNP data with a special reference to the Ainu and the Ryukyuan populations. *J. Hum. Genet.* **57**, 787–795 (2012).
22. Wollstein, A. et al. Demographic history of Oceania inferred from genome-wide data. *Curr. Biol.* **20**, 1983–1992 (2010).
23. Kayser, M. et al. Genome-wide analysis indicates more Asian than Melanesian ancestry of Polynesians. *Am. J. Hum. Genet.* **82**, 194–198 (2008).
24. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
25. Raghavan, M. et al. The genetic prehistory of the New World Arctic. *Science* **345**, 1255832 (2014).
26. Brace, C. L. et al. Old World sources of the first New World human inhabitants: a comparative craniofacial view. *Proc. Natl Acad. Sci. USA* **98**, 10017–10022 (2001).
27. Jantz, R. L. & Owsley, D. W. Variation among early North American crania. *Am. J. Phys. Anthropol.* **114**, 146–155 (2001).
28. Jantz, R. L. & Owsley, D. W. in *Paleoamerican Origins: Beyond Clovis* (eds Bonnichen, R., Lepper, B., Stanford, D. J. & Waters, M. R.) 267–275 (Texas A & M Univ. Press, 2005).
29. Howells, W. W. Howells' craniometric data on the internet. *Am. J. Phys. Anthropol.* **101**, 441–442 (1996).
30. Lawrence, D. M. et al. Mitochondrial DNA of protohistoric remains of an Arikara population from South Dakota: implications for the macro-Siouan language hypothesis. *Hum. Biol.* **82**, 157–178 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the members of the Confederated Tribes of the Colville Reservation who provided their DNA for this study. We would like to also thank Nez Perce Tribe, Umatilla Tribe, Yakama Nation, and Wanapum Band for their consultation. We thank C. B. Pulliam and G. C. Celmer of the US Army Corps of Engineers, and the Burke Museum of the University of Washington, especially L. Phillips. We thank the Danish National Research Foundation and the Lundbeck Foundation for financial support and the Danish National Sequencing Centre for help with sequencing. M.R. was funded by grant 12-131829 from the Danish Council for Independent Research. J.V.M.-M. was supported by 'Consejo Nacional de Ciencia y Tecnología' (Mexico). G.D.P. was supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1147470. I.M. was funded by a DFF-YDUN grant from the Danish Council for Independent Research. C.V. was funded by Marie Curie Intra-European Fellowship-FP7-People-PIEF-GA-2009-255503

Author Contributions E.W. headed the project. E.W. and T.W.S. initiated the work. D.J.M., R.S.M. and C.V. provided archaeological knowledge and ethical advice. M.R. conducted the laboratory work. M.R., M.S., A.A., T.S.K., J.V.M.-M., G.D.P., M.E.A., I.M., H.J., L.O., C.D.B. and R.N. conducted the DNA analyses. C.P.E.Z. and M.S.P.L. conducted the morphological analyses. E.W., D.J.M. and M.R. wrote the paper with contributions from all the authors.

Author Information The Kennewick Man genome sequence has been deposited in the Sequence Read Archive under accession number SRS937952. The Colville SNP chip data can be used for scientific confirmation of our findings (with regard to the group's direct ancestry to Kennewick Man). It cannot be used for any other purpose. The Colville data in this study are available under data access agreement with E.W. The bam file with mapped data can be downloaded from (<http://purl.stanford.edu/bg616nn8691>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.W. (ewillerslev@snm.ku.dk).

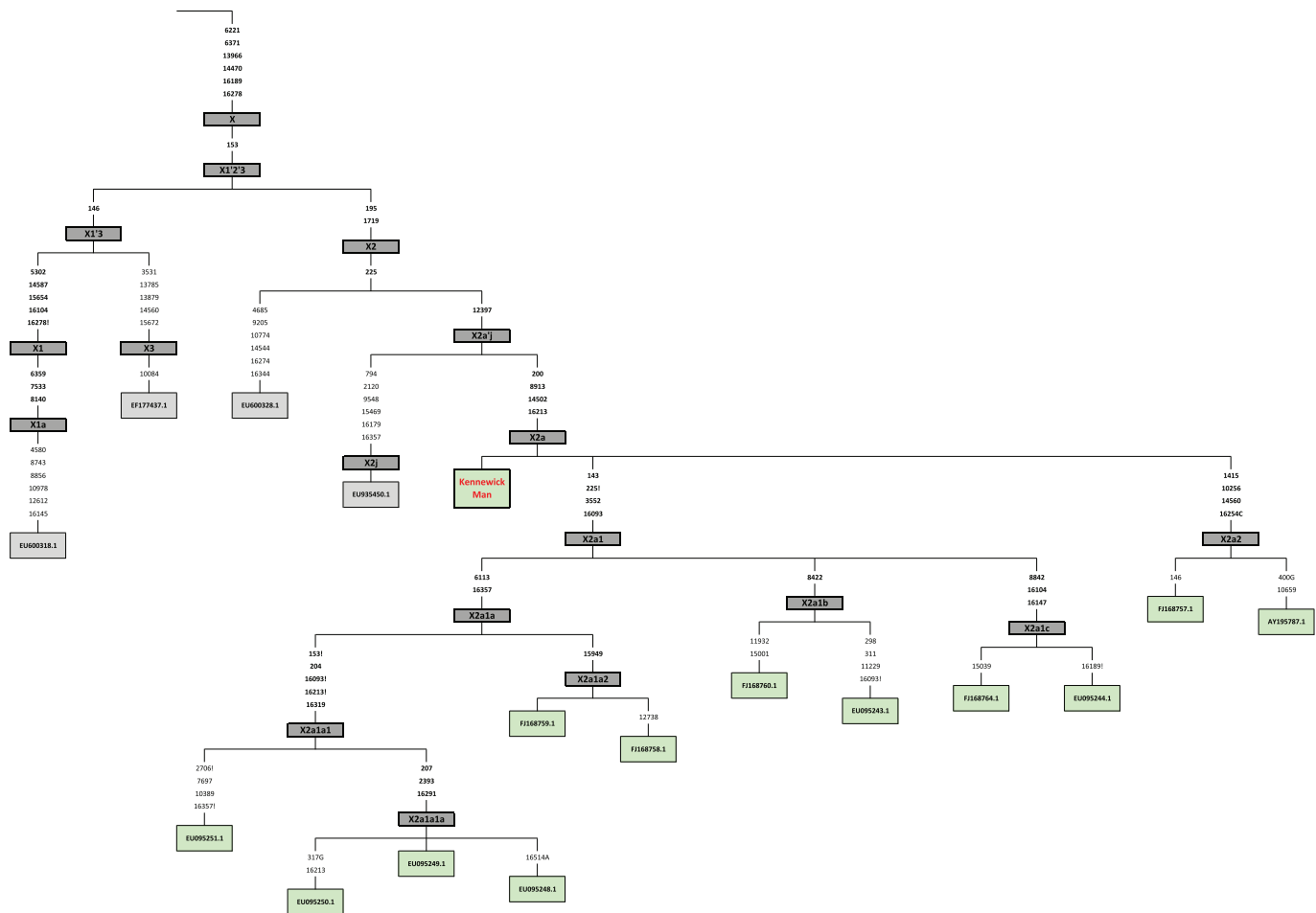
 This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

METHODS

We extracted DNA from a 200-mg bone fragment from Kennewick Man, and built both single and double stranded DNA libraries, which were sequenced on the Illumina HiSeq platform (Supplementary Information sections 1, 2). We performed DNA damage analyses and estimated decay rates to verify authenticity; additionally we estimated contamination on both nuclear and mitochondrial DNA (Supplementary Information sections 2, 3 and 4). For the nuclear contamination we developed a model to identify the most likely source population (Supplementary Information section 4). Both mitochondrial and Y-chromosome haplogroup were determined (Supplementary Information sections 2 and 5). To resolve the ancestry of Kennewick Man, we performed PCA, outgroup f_3 - and D -statistics, as well as ADMIXTURE analyses on a panel of published SNP array data that was collected and curated from worldwide populations with suggested relationship to Kennewick Man (Supplementary Information sections 6 and 7), in addition to data generated from members of the Colville Tribe (Supplementary Information section 1). Individual and tribal consent was obtained for all study participants, and the National Committee on Health Research Ethics in Denmark had no comments

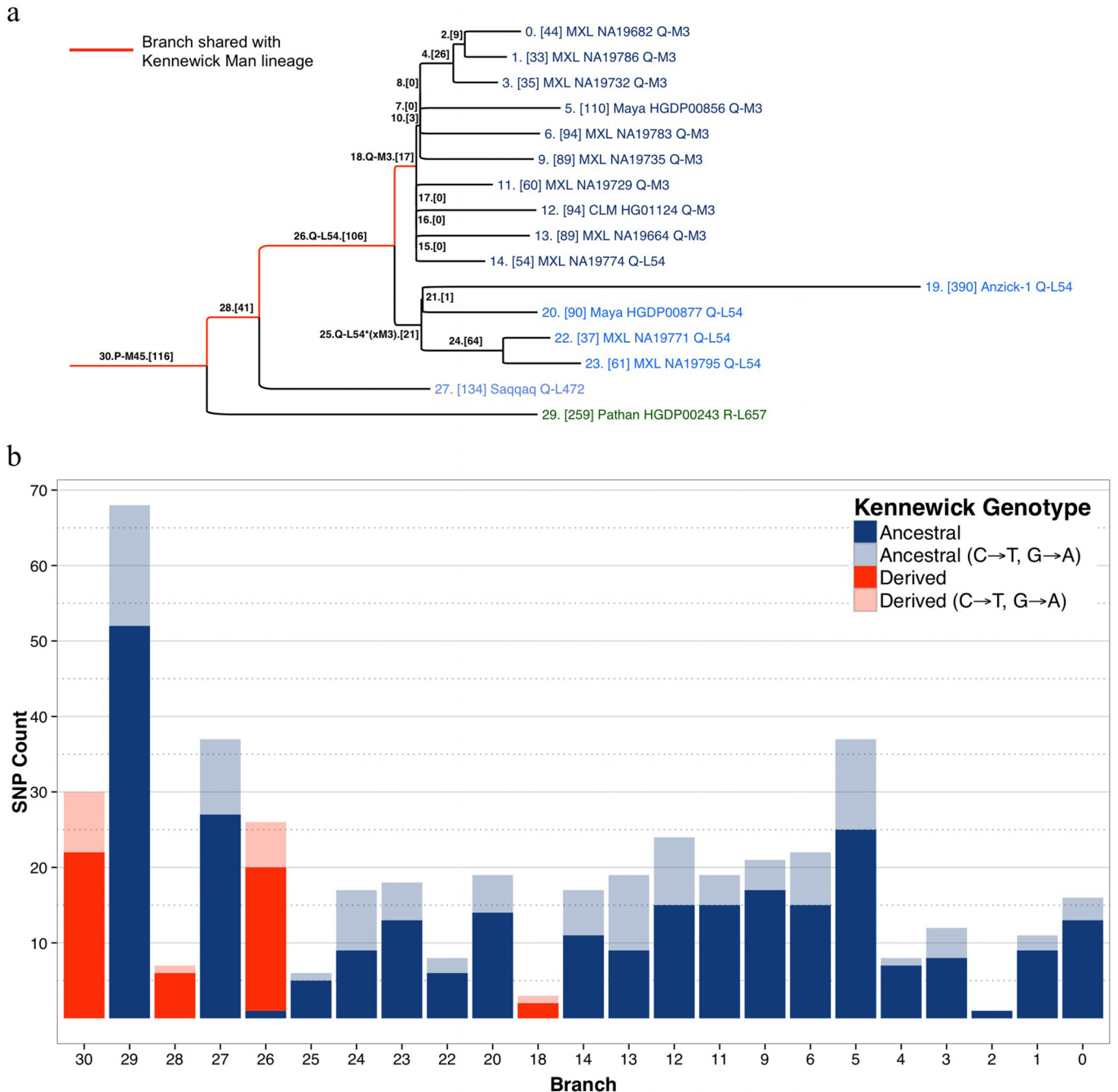
on the design (H-3-2012-FSP21). We tested if Kennewick Man belonged to a population ancestral to the Colville Tribe and estimated their divergence time (Supplementary Information section 8). Lastly, we reanalysed the craniometric data for Kennewick Man, and compared it to both individual samples and population mean data (Supplementary Information section 9).

Population labels: agg, Algonquin; ain, Ainu; ale, Aleutian; alt, Altai; arh, Arhuaco; aym, Aymara; bur, Buryat; cab, Cabecar; ceu/CEU, Utah residents with ancestry from northern and western Europe; chb/CHB, Han Chinese in Beijing, China; chi, Chipewyan; chl, Chilote; chu, Chukchi; cvi, Colville; dia, Diaguita; dol, Dolgan; eGl, EastGreenland; emb, Embera; eve, Even; evk, Evenki; ghb, Guahibo; gua, Guarani; guy, Guaymi; hai, Haida; ing, Inga; kaq, Kaqchikel; kar, Karitiana; kha, Khakas; kor, Koryak; mix, Mixe; mon, Mongol; mxt, Mixtec; my1, Maya1; my2, Maya2; nga, Nganasan; nsi, Nisgaa; oji, Ojibwa; pia, Piapoco; pim, Pima; pol, Polynesia; que, Quechua; sel, Selkup; spl, Splatkin; sts, Stswecem'c; sur, Surui; tep, Tepehuano; tic, Ticuna; tli, Tlingit; tsi, Tsimshian; tuv, Tuvanian; wGl, WestGreenland; way, Wayuu; wic, Wichi; yri/YRI, Yoruba in Ibadan, Nigeria; za1, Zapotec1; za2, Zapotec2.



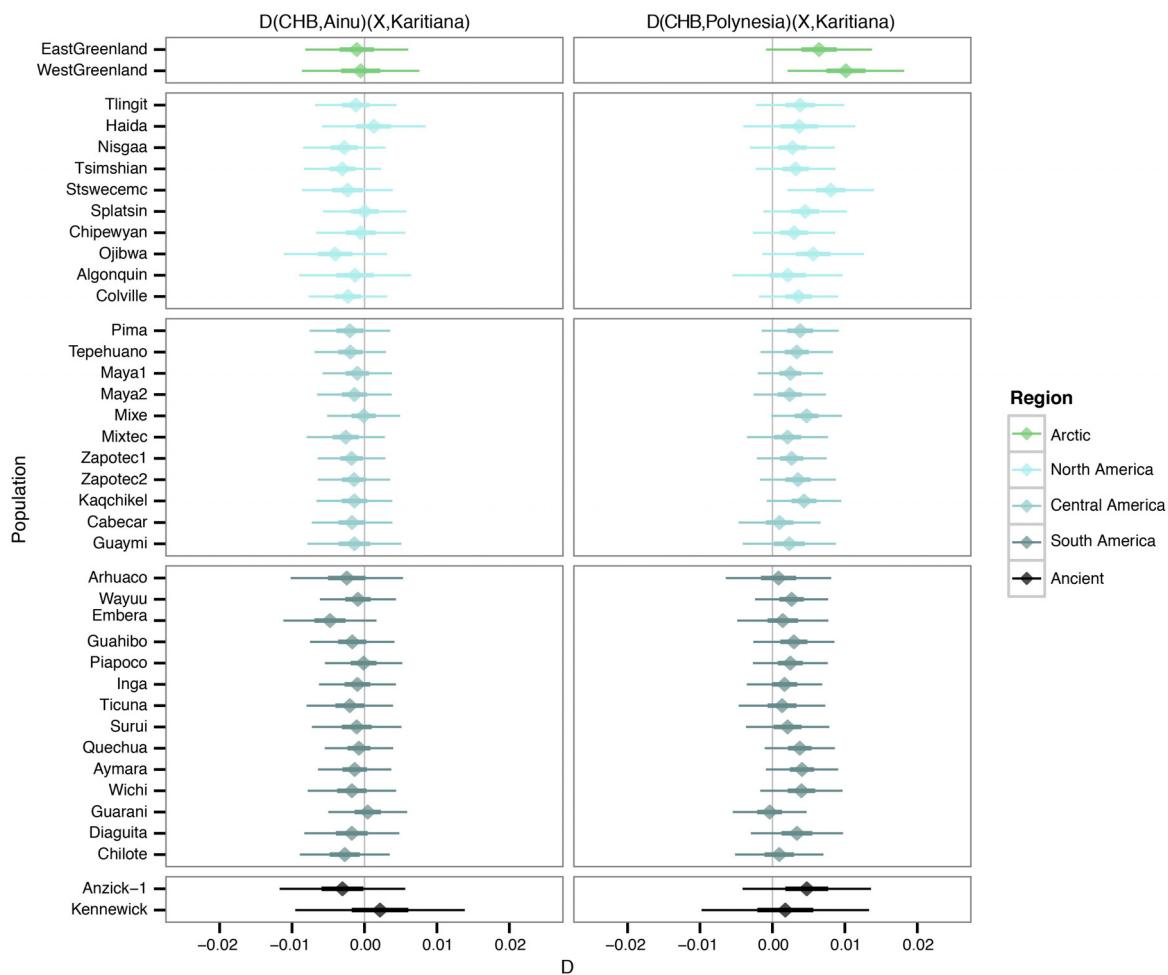
Extended Data Figure 1 | Phylogenetic tree of mitochondrial haplogroup X including Kennewick Man. A median-joining network of GenBank sequences from haplogroup X was constructed as described in the Supplementary Information. Haplogroup names are indicated by bold dark grey boxes,

sequences of Native American origin are in light green background. Back mutations to ancestral state are denoted with an ! symbol. GenBank accession numbers are shown in boxes at branch tips.



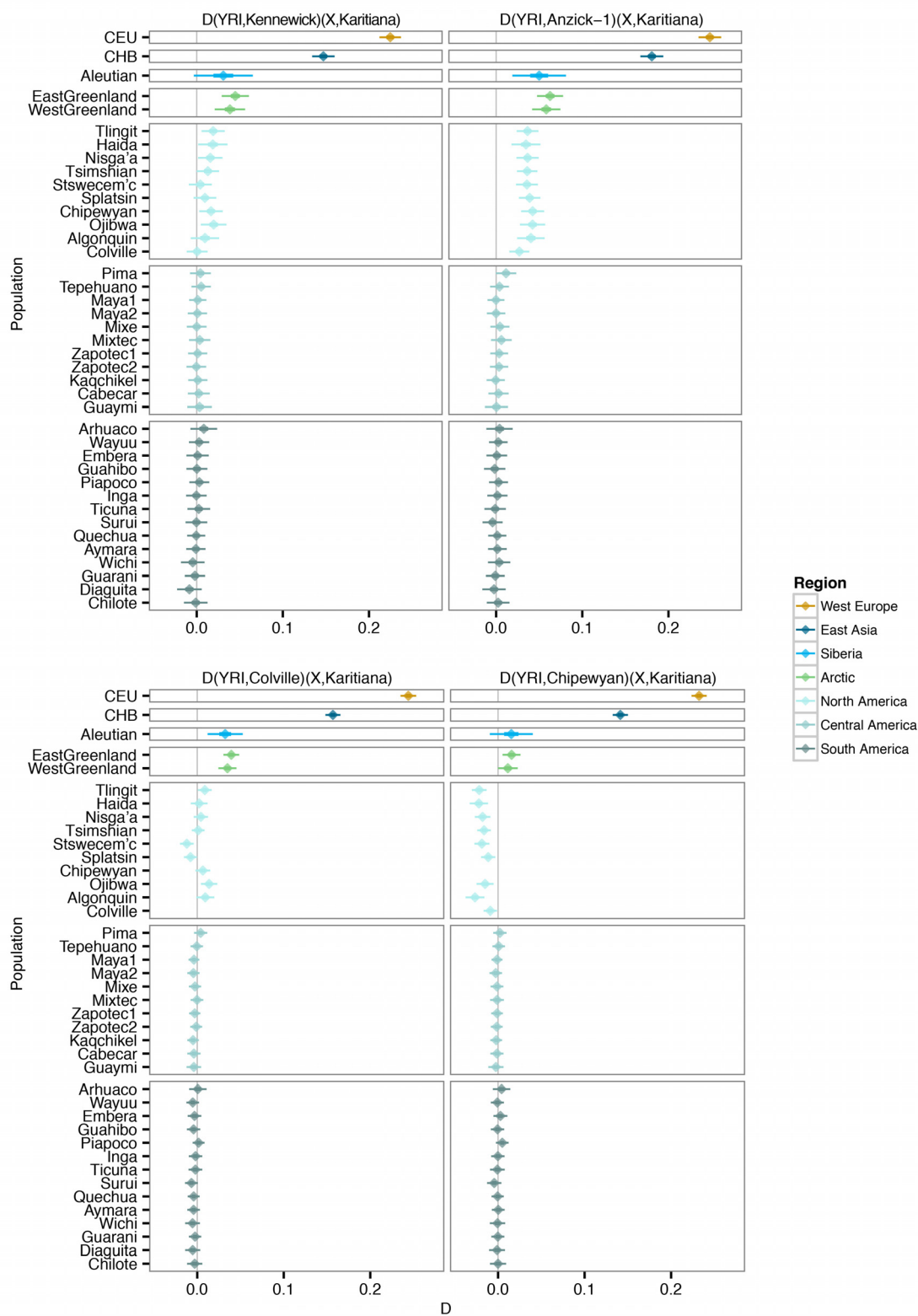
Extended Data Figure 2 | Y-chromosome haplogroup. **a**, Phylogenetic tree including representative sequences of haplogroup P, the clade that includes haplogroups Q and R. Kennewick Man shares ancestry with orange branches. Each branch is labelled with an integer index and, in brackets, the number of SNPs that define it. **b**, Counts of SNPs from each branch of the tree, stratified by

Kennewick Man genotype (ancestral in blue, derived in orange) and mutation type (C→T and G→A transitions coloured more lightly). Branch 19 was omitted to preserve scale; the Kennewick genotype was ancestral at all 145 sites for which read data were available.



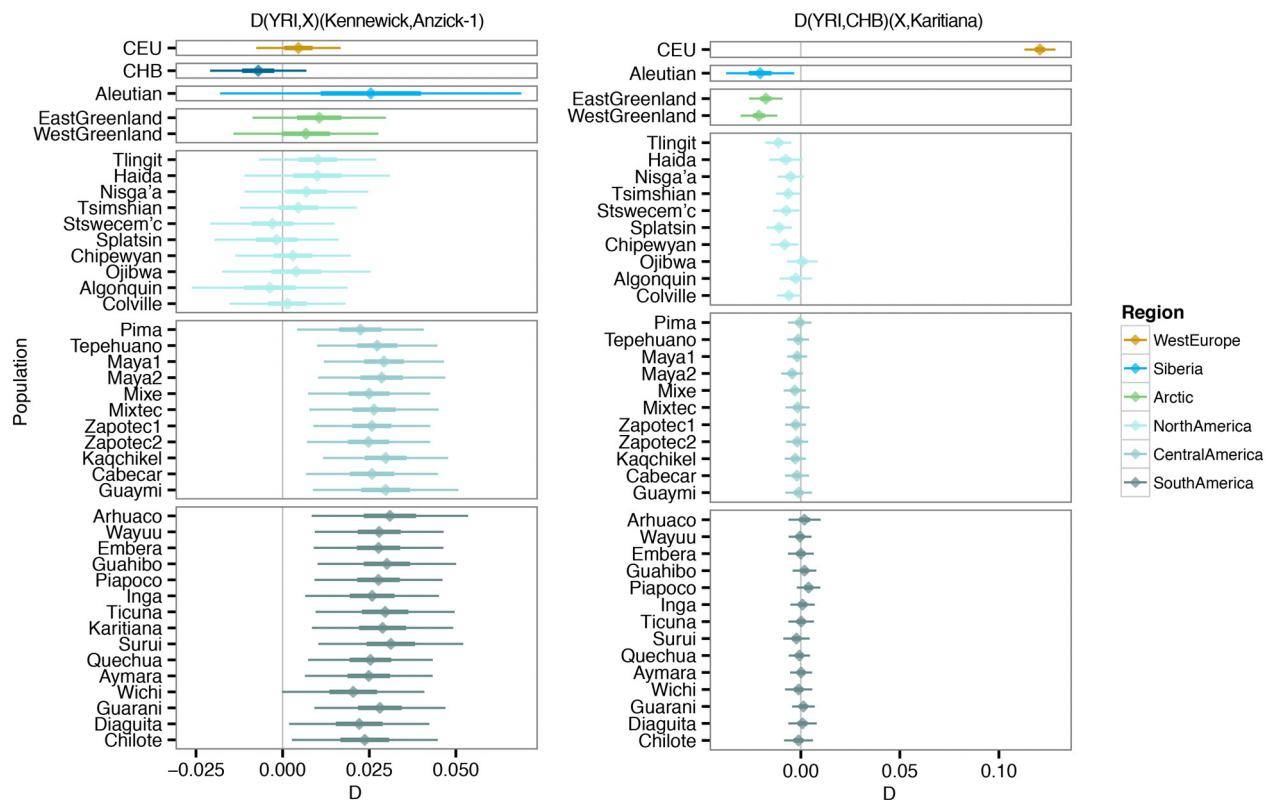
Extended Data Figure 3 | D-statistic-based test for Ainus and Polynesian affinity to Han Chinese and Native Americans. Test of the type $D((CHB,Ainu/Polynesian),(X,Karitiana))$, where X is an Arctic or Native American population, including Kennewick Man and Anzick-1. Values equal

to 0, has population X being closer to Karitiana than either CHB or Ainus/Polynesian. Thick and thin whiskers represent 1 and 3 standard errors, respectively.



Extended Data Figure 4 | Similarity between Kennewick Man and Anzick-1 as well as Colville. Test of $D((YRI, Kennewick/Anzick-1/Colville/Chipewyan), (X, Karitiana))$, to illustrate similarities between Kennewick Man

and Anzick-1, and the Colville tribe. Thick and thin whiskers represent 1 and 3 standard errors, respectively.



Extended Data Figure 5 | Affinity of Kennewick Man and Anzick-1 to Native American populations. Test of $D((YRI,X),(Kennewick,Anzick-1))$ to distinguish different affinities of the ancient samples, and

$D((YRI,CHB),(X,Karitiana))$ to test for recent Asian gene flow. Thick and thin whiskers represent 1 and 3 standard errors, respectively.

Directional dominance on stature and cognition in diverse human populations

A list of authors and their affiliations appears in the online version of the paper

Homozygosity has long been associated with rare, often devastating, Mendelian disorders¹, and Darwin was one of the first to recognize that inbreeding reduces evolutionary fitness². However, the effect of the more distant parental relatedness that is common in modern human populations is less well understood. Genomic data now allow us to investigate the effects of homozygosity on traits of public health importance by observing contiguous homozygous segments (runs of homozygosity), which are inferred to be homozygous along their complete length. Given the low levels of genome-wide homozygosity prevalent in most human populations, information is required on very large numbers of people to provide sufficient power^{3,4}. Here we use runs of homozygosity to study 16 health-related quantitative traits in 354,224 individuals from 102 cohorts, and find statistically significant associations between summed runs of homozygosity and four complex traits: height, forced expiratory lung volume in one second, general cognitive ability and educational attainment ($P < 1 \times 10^{-300}$, 2.1×10^{-6} , 2.5×10^{-10} and 1.8×10^{-10} , respectively). In each case, increased homozygosity was associated with decreased trait value, equivalent to the offspring of first cousins being 1.2 cm shorter and having 10 months' less education. Similar effect sizes were found across four continental groups and populations with different degrees of genome-wide homozygosity, providing evidence that homozygosity, rather than confounding, directly contributes to phenotypic variance. Contrary to earlier reports in substantially smaller samples^{5,6}, no evidence was seen of an influence of genome-wide homozygosity on blood pressure and low density lipoprotein cholesterol, or ten other cardio-metabolic traits. Since directional dominance is predicted for traits under directional evolutionary selection⁷, this study provides evidence that increased stature and cognitive function have been positively selected in human evolution, whereas many important risk factors for late-onset complex diseases may not have been.

Inbreeding influences complex traits through increases in homozygosity and corresponding reductions in heterozygosity, most likely resulting from the action of deleterious (partially) recessive mutations⁸. For polygenic traits, a systematic association with genome-wide homozygosity is not expected when dominant alleles at some loci increase the trait value while others decrease it. Rather, dominance must be biased in one direction on average over all causal loci, for instance to decrease the trait. Such directional dominance is expected to arise in evolutionary fitness-related traits due to directional selection⁸. Studies of genome-wide homozygosity thus have the potential to reveal the non-additive allelic architecture of a trait and its evolutionary history. Historically, inbreeding has been measured using pedigrees⁹. However, such techniques cannot account for the stochastic nature of inheritance, nor are they practical for the capture of the distant parental relatedness present in most modern-day populations. High-density genome-wide single nucleotide polymorphism (SNP) array data can now be used to assess genome-wide homozygosity directly, using genomic runs of homozygosity (ROH). Such runs are inferred to be homozygous-by-descent and are common in human populations^{10,11}. Summed ROH (SROH) is the sum of the length of these ROH, in megabases of DNA. F_{ROH} is the ratio of SROH to the

total length of the genome. Like pedigree-based F (with which it is highly correlated³), F_{ROH} estimates the probability of being homozygous at any site in the genome. F_{ROH} has been shown to vary widely within and between populations¹² and is a powerful method of detecting genome-wide homozygosity effects¹³.

We found marked differences by geography and demographic history in both the population mean SROH and the relationship between SROH and NROH (the numbers of separate runs of homozygosity) (Fig. 1). As observed previously^{3,12,14}, isolated populations have a higher burden of ROH, whereas African heritage populations have the least homozygosity.

We studied $\beta_{F_{\text{ROH}}}$, defined as the effect of F_{ROH} on 16 complex traits of biomedical importance (Fig. 2). For height, FEV1 (forced expiratory volume in one second, a measure of lung function), educational attainment, and g (a measure of general cognitive ability derived from scores on several diverse cognitive tests), we found the effect sizes were greater than two intra-sex standard deviations, with P values all less than 10^{-5} . Thus the associations could not plausibly be explained by chance alone (Table 1; see Extended Data Figs 1–4 for Forest plots of individual traits; Supplementary Table 1 for s.d.). To ensure that the results were not driven by a few outliers, we repeated the analysis excluding extreme sub-cohort trait results. In all cases the effect sizes and their significance remained similar or increased (see Supplementary Table 2 for comparisons with and without outliers). After exclusion of outliers, these effect sizes translate into a reduction of 1.2 cm in height and 137 ml in FEV1 for the offspring of first cousins, and into a decrease of 0.3 s.d. in g and 10 months' less educational attainment.

We performed a number of analyses to exclude confounding. While SROH is wholly a genetic effect, its inheritance is entirely non-additive. Therefore, unlike in genome-wide association, an association with population genetic structure or co-segregation of additive genome-wide polygenic effects and SROH (as opposed to SNPs in a genome-wide association study) are not expected as a matter of course, except in the case of siblings. However, confounding could still theoretically arise, as discussed below. We therefore assessed this by conducting stratified and covariate analyses. We found effects of similar magnitude and in the same direction for all four traits across isolated and non-isolated European, Finnish, African, Hispanic, East Asian and South and Central Asian populations (Extended Data Fig. 5a and Supplementary Table 3). We further tested whether the effect sizes were similar when cohorts were split into more and less homozygous groups. The effect sizes were very similar, even though the degree of homozygosity (and variation in homozygosity) varied 3–10-fold between the two strata (depending on which cohorts contributed to the trait; Extended Data Fig. 5b). This suggests a broadly linear relationship with SROH. In general, confidence intervals overlap for stratified estimates, suggesting that differences arose due to sampling variance. Larger confidence intervals for some estimates reflect the lower power of some strata, in turn reflecting the sample size and degree of homozygosity of those strata (for example, the wider confidence intervals for estimates of educational attainment $\beta_{F_{\text{ROH}}}$ for Finnish and African strata). Finally, we fitted educational attainment

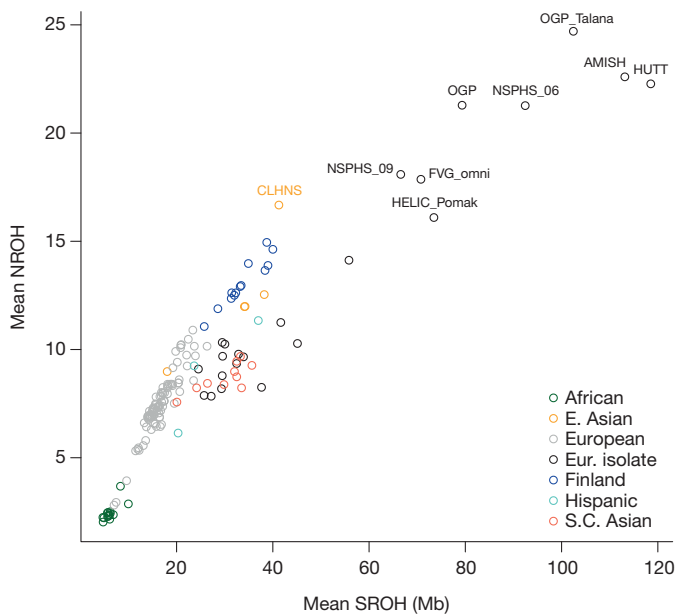


Figure 1 | Runs of homozygosity by cohort. The sum of runs of homozygosity (SROH) and the number of runs of homozygosity (NROH) are shown by sub-cohort. Populations differ by an order of magnitude in their mean burden of ROH. There are clear differences by continent and population type both in the mean SROH, and the relationship between SROH and NROH. S.C. Asian, South and Central Asian; E. Asian, East Asian; Eur. Isolate, European isolates. The ten most homozygous cohorts are labelled: AMISH, Old Order Amish from Lancaster County, Pennsylvania; HUTT, Schmiedeleut Hutterites from South Dakota; NSPHS, northern Swedish population health study, 06 and 09 suffixes are different sampling years from different counties in northern Sweden; OGP, Ogliastro genetic park, Sardinia, Italy; Talana is a particular village in the region; FVG, Friuli-Venezia-Giulia genetic park, Italy, omni and 370 suffixes refer to subsets genotyped with the Illumina OmniX and 370CNV arrays; HELIC, Hellenic isolates, Greece, from Pomak villages in Thrace, and CLHNS, Cebu Longitudinal Health and Nutrition Survey in the Philippines.

as a proxy for potential confounding by socio-economic status; this covariate was available in sufficient (47) cohorts to maintain power. The estimated effect sizes for height, FEV1 and g all reduced (17%, 18% and 35%, respectively, Extended Data Fig. 5c), but this might have been expected given the known covariance between these three traits and educational attainment, and the association we found between educational attainment and F_{ROH} . We found very small differences (3–11% reductions) in estimated $\beta_{F_{ROH}}$ (Extended Data Fig. 6 and Supplementary Table 4), when comparing the fitting of polygenic mixed models as opposed to fixed-effect-only models, again suggesting that confounding (in this case due to polygenic effects arising from recent common ancestry) did not substantially affect the results.

Despite the observed 17–35% reductions in estimated effect sizes for F_{ROH} on height, FEV1 and g , when fitting educational attainment as a covariate, the persistence of an effect suggests that most of the signals we observe are genetic. The consistency of effects with and without fitting relatedness and in particular in populations with very different degrees of homozygosity, all appear inconsistent with confounding as a result of environmental or additive genetic effects. As does the broad similarity in effect sizes across continents, although the relatively smaller numbers of cohorts of non-European descent meant we had limited power to detect intercontinental differences in effect sizes.

It is also interesting to consider the potential influence of assortative mating, which is commonly observed for human stature, cognition and education. The phenotypic extremes could be more genetically similar to each other and hence the offspring more homozygous, even if the highly polygenic trait architectures reduce this effect. However, at

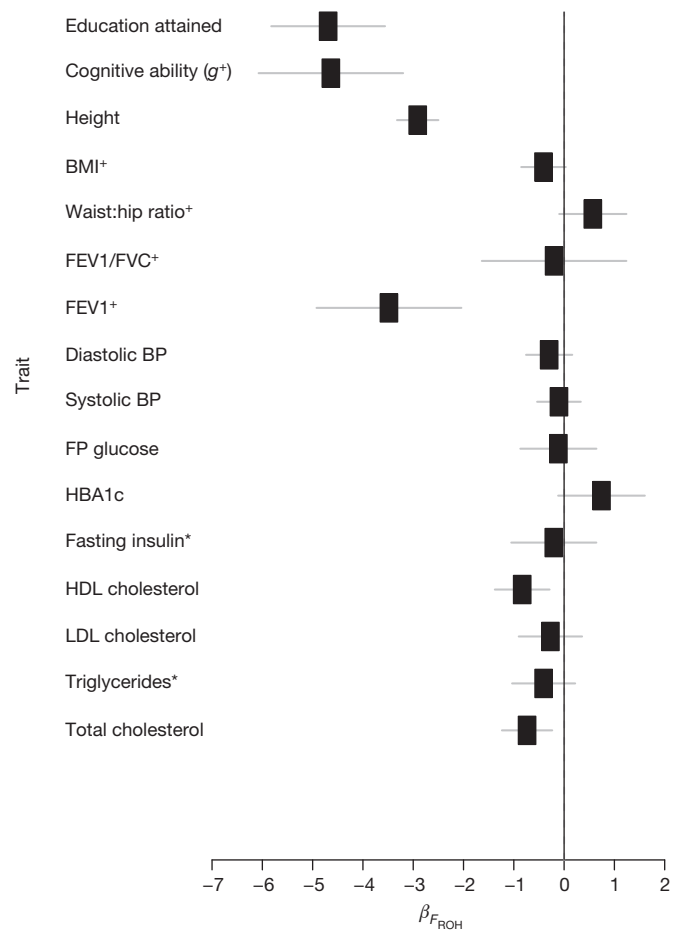


Figure 2 | Effects of genome-wide homozygosity, $\beta_{F_{ROH}}$, on 16 traits. Four phenotypes show a significant effect of burden of ROH: height (145 sub-cohorts), FEV1 (34), educational attainment (47) and general cognitive ability, g (23). HDL and total cholesterol are not significantly different from zero after correcting for 16 tests and no effect is observed for the other traits. To account for the different numbers of males and females in cohorts and marked effect of sex on some traits, trait units are intra-sex standard deviations. $\beta_{F_{ROH}}$ is the estimated effect of F_{ROH} on the trait, where F_{ROH} is the ratio of the SROH to the total length of the genome. 95% confidence intervals are also plotted. Plus signs indicate that the phenotype was rank transformed, asterisks indicate that the phenotype was log transformed. BMI, body mass index; BP, blood pressure; FP fasting plasma; HbA1c, haemoglobin A1c (glycated haemoglobin); FEV1, forced expiratory volume in one second; FVC, forced vital capacity; HDL, high density lipoprotein; LDL, low density lipoprotein.

least in its simplest balanced form, the increase in genetic similarity would be equal at both ends of the phenotypic distribution, leading to no linear association between such genetic similarity and the trait; both tall and short people would be more homozygous. Furthermore, humans also mate assortatively on the basis of body mass index, for which we see no effect. A more complex possibility, a form of reverse causality, could arise when subjects from one trait extreme (for example, people with high educational attainment) are on average more geographically mobile, and thus have less homozygous offspring, with those offspring in turn inheriting the trait extreme concerned¹⁵. We do not think that this mechanism can account for our results, since it does not readily explain the constancy of our results under different models, especially the similarity in $\beta_{F_{ROH}}$ for either more or less homozygous populations. Moreover, we observe similar effects in multiple single-village cohorts, and the Amish and Hutterites, where there is no geographic structure and/or no sampling of immigrants, hence such confounding by differential migration cannot occur.

Table 1 | Effects of genome-wide burden of runs of homozygosity on four traits

Phenotype	Outliers	Height	FEV1 ⁺	Educational attainment	Cognitive ability (g) ⁺
Subjects		354,224	64,446	84,725	53,300
P-association	Included	$<1 \times 10^{-300}$	2.1×10^{-6}	1.8×10^{-10}	2.5×10^{-10}
P-heterogeneity	Included	0.014	0.10	1.2×10^{-5}	0.071
$\beta_{F_{ROH}}$-s.d.	Excluded	-2.91	-3.48	-4.69	-4.64
s.e. $\beta_{F_{ROH}}$-s.d.	Excluded	0.21	0.73	0.58	0.73
$\beta_{F_{ROH}}$-units	Excluded	-0.188	-2.2	-12.9	-4.64
s.e. $\beta_{F_{ROH}}$-units	Excluded	0.014	0.46	1.83	0.73
Units		m	litres	years	s.d.
First cousin offspring effect	Excluded	-1.2	-137	-9.7	-0.29
Units		cm	ml	months	s.d.

P-association is the *P* value for association; P-heterogeneity is the *P* value for heterogeneity in a meta-analysis between trait and unpruned F_{ROH} ; $\beta_{F_{ROH}}$ -s.d. is the effect size estimate of F_{ROH} expressed in units of intra-sex phenotypic standard deviations; $\beta_{F_{ROH}}$ -units is the effect size estimate for $F_{ROH} = 1$ expressed in the measurement units; s.e., standard error. The *P* values for those traits showing evidence for association are calculated, including five outlying cohort-specific effect size estimates (an outlier was defined as *t*-test statistic over 3 for the null hypothesis that the cohort effect size estimate equals the meta-analysis effect size estimate), which is conservative, as the majority of these are in the opposite direction. However, β estimates exclude these outliers, for which there is evidence of discrepancy, and should thus be more accurate. A plus symbol indicates that the phenotype was rank transformed; FEV1 is forced expiratory lung volume in one second; general cognitive ability is calculated as the first unrotated principal component of test scores across diverse domains of cognition.

Our estimate for the effect of homozygosity in height is consistent with previous work: genomic⁴ and pedigree¹⁶ studies have shown genome-wide homozygosity effects on stature with similar effect sizes (a 0.01 increase in *F* decreases height by 0.037 s.d.¹⁶ versus 0.029 s.d. in the present study). We speculate that homozygosity is acting on a shared endophenotype of torso size which we detect in the height and FEV1 traits. The fact that the FEV1:FVC (forced vital capacity) ratio is not associated with ROH points to the effect acting on lung/chest size rather than airway calibre. The cognition effects cannot be wholly generated by height as an intermediate cause, given the greater proportion of variance explained for cognition, although we note that the correlation between height and cognition has been estimated as 0.16 (standard error, s.e. = 0.01), and the genetic correlation (the correlation in additive genetic values) as 0.28 (s.e. = 0.09; ref. 17).

Height is the canonical human complex trait, highly heritable and polygenic, with 697 genome-wide significant variants in 423 loci explaining 20% of the heritability and all common variants predicted to explain 60% of the heritability¹⁸. Most of the genetic architecture appears to be additive in nature, however ROH analysis reveals a distinct directional dominance component.

Our genomic confirmation of directional dominance for *g* and discovery of genome-wide homozygosity effects on educational attainment in a wide range of human populations adds to our knowledge of the genetic underpinnings of cognitive differences, which are currently thought to be largely due to additive genetic effects¹⁹. Our findings go beyond earlier pedigree-based analyses of recent consanguinity to demonstrate that the observed effect of genome-wide homozygosity is not a result of confounding and influences demographically diverse populations across the globe. The estimated effect size is consistent with pedigree data (a 0.01 increase in *F* decreases *g* by 0.046 s.d. in our analysis and 0.029–0.048 s.d. in pedigree-based studies)²⁰. It is germane to note that one extreme of cognitive function, early onset cognitive impairment, is strongly influenced by deleterious recessive loci²¹, so we can speculate that an accumulation of recessive variants of weaker effect may influence normal variation in cognitive function. Although increasing migration and panmixia have generated a secular trend in decreasing homozygosity²², the Flynn effect, wherein succeeding generations perform better on cognitive tests than their predecessors²³, this cannot be explained by our findings, because the intergenerational change in cognitive scores is much larger than the differences in homozygosity would predict. Likewise, the genome-wide homozygosity effect on height cannot explain a significant proportion of the observed intergenerational increases²⁴.

Inbreeding depression, which arises from the effect of genome-wide homozygosity, is ubiquitous in plants and is seen for numerous fitness-related traits in animals²⁵, but we observed no effect for the 12 other mainly cardio-metabolic traits in which variation is strongly related to age. This suggests that previous reports in ecological studies or

substantively smaller studies using pedigrees or relatively small numbers of genetic markers may have been false positives^{5,6}. The lack of directional dominance on these traits does not, however, rule out a recessive component, as recessive variants acting in different directions will cancel out. Dominance variance is predicted to be greater for late-onset fitness traits²⁶, so the lack of genome-wide homozygosity effects in the cardio-metabolic traits may be due to lack of directional dominance. ROH analyses within specific genomic regions are warranted to map recessive effects even when there is no genome-wide directional dominance. Such recessive effects have been observed for a subset of cardiovascular risk factors²⁷ and expression traits²⁸.

We have demonstrated the existence of directional dominance on four complex traits (stature, lung function, cognitive ability and educational attainment), while showing any effect on another 12 health-related traits is at least almost an order of magnitude smaller, non-linear or non-existent. This directional dominance implies that size and cognition (like schizophrenia protective alleles²⁹) have been positively selected in human history – or at least that some variants increasing these traits contribute to fitness. However, the lack of any evidence for an association between many late-onset cardiovascular disease risk factors and ROH is perhaps surprising and suggests testing directly for an association between ROH and disease outcome. The magnitude of genome-wide homozygosity effects is relatively small in all cases, thus Darwin's supposition³⁰ of “any evil [of inbreeding] being very small” is substantiated.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 February; accepted 28 May 2015.

Published online 1 July 2015.

- Garrod, A. The incidence of alkaptonuria: a study of chemical individuality. *Lancet* **160**, 1616–1620 (1902).
- Darwin, C. *The Variation of Animals and Plants Under Domestication* (Appleton, 1868).
- McQuillan, R. *et al.* Runs of homozygosity in European populations. *Am. J. Hum. Genet.* **83**, 359–372 (2008).
- McQuillan, R. *et al.* Evidence of inbreeding depression on human height. *PLoS Genet.* **8**, e1002655 (2012).
- Rudan, I. *et al.* Inbreeding and the genetic complexity of human hypertension. *Genetics* **163**, 1011–1021 (2003).
- Campbell, H. *et al.* Effects of genome-wide heterozygosity on a range of biomedically relevant human quantitative traits. *Hum. Mol. Genet.* **16**, 233–241 (2007).
- Charlesworth, D. & Willis, J. H. The genetics of inbreeding depression. *Nature Rev. Genet.* **10**, 783–796 (2009).
- Wright, S. *Evolution and the Genetics of Populations, Vol. 3: Experimental Results and Evolutionary Deductions* (University of Chicago Press, 1977).
- Wright, S. Coefficients of inbreeding and relationships. *Am. Nat.* **56**, 330–339 (1922).
- Broman, K. W. & Weber, J. L. Long homozygous chromosomal segments in reference families from the Centre d'Étude du Polymorphisme Humain. *Am. J. Hum. Genet.* **65**, 1493–1500 (1999).

11. Gibson, J., Morton, N. E. & Collins, A. Extended tracts of homozygosity in outbred human populations. *Hum. Mol. Genet.* **15**, 789–795 (2006).
12. Kirin, M. *et al.* Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE* **5**, e13996 (2010).
13. Keller, M. C., Visscher, P. M. & Goddard, M. E. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* **189**, 237–249 (2011).
14. Pemberton, T. J. & Rosenberg, N. A. Population-genetic influences on genomic estimates of the inbreeding coefficient: a global perspective. *Hum. Hered.* **77**, 37–48 (2014).
15. Abdellaoui, A. *et al.* Educational attainment influences levels of homozygosity through migration and assortative mating. *PLoS ONE* **10**, e0118935 (2015).
16. Neel, J. V. *et al.* The effects of parental consanguinity and inbreeding in Hirado, Japan. II. Physical development, tapping rate, blood pressure, intelligence quotient, and school performance. *Am. J. Hum. Genet.* **22**, 263–286 (1970).
17. Marioni, R. E. *et al.* Common genetic variants explain the majority of the correlation between height and intelligence: the generation Scotland study. *Behav. Genet.* **44**, 91–96 (2014).
18. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genet.* **46**, 1173–1186 (2014).
19. Deary, I. J. *et al.* Genetic contributions to stability and change in intelligence from childhood to old age. *Nature* **482**, 212–215 (2012).
20. Morton, N. E. Effect of inbreeding on IQ and mental retardation. *Proc. Natl Acad. Sci. USA* **75**, 3906–3908 (1978).
21. Najmabadi, H. *et al.* Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* **478**, 57–63 (2011).
22. Nalls, M. A. *et al.* Measures of autozygosity in decline: globalization, urbanization, and its implications for medical genetics. *PLoS Genet.* **5**, e1000415 (2009).
23. Flynn, J. R. Massive IQ gains in 14 nations: what IQ tests really measure. *Psychol. Bull.* **101**, 171–191 (1987).
24. Galton, F. *Natural inheritance* (MacMillan, 1889).
25. Hoffman, J. I. *et al.* High-throughput sequencing reveals inbreeding depression in a natural population. *Proc. Natl Acad. Sci. USA* **111**, 3775–3780 (2014).
26. Wright, A., Charlesworth, B., Rudan, I., Carothers, A. & Campbell, H. A polygenic basis for late-onset disease. *Trends Genet.* **19**, 97–106 (2003).
27. Weiss, L. A., Pan, L., Abney, M. & Ober, C. The sex-specific genetic architecture of quantitative traits in humans. *Nature Genet.* **38**, 218–222 (2006).
28. Powell, J. E. *et al.* Congruence of additive and non-additive effects on gene expression estimated from pedigree and SNP data. *PLoS Genet.* **9**, e1003502 (2013).
29. Keller, M. C. *et al.* Runs of homozygosity implicate autozygosity as a schizophrenia risk factor. *PLoS Genet.* **8**, e1002656 (2012).
30. Darwin, C. *The Effects of Crossing and Self Fertilization in the Vegetable Kingdom* (John Murray, 1876).

Supplementary Information is available in the online version of the paper.

Acknowledgements This paper is the work of the ROHgen consortium. We thank the participants in all ROHgen studies; cohort-specific acknowledgements are detailed in Supplementary Table 6. This work was funded by a UK Medical Research Council (MRC) PhD studentship to P.K.J.; and J.F.W. and O.P. acknowledge support from the MRC Human Genetics Unit “QTL in Health and Disease” programme. We thank W. G. Hill for discussions and comments on the manuscript and K. Lindsay for administrative assistance.

Author contributions C.Hal., P.N., M.Me., H.B., N.J.S., D.C., D.A.M., R.S.C., P.F., G.P., S.F.G., H.H., L.F., R.A.S., A.D.M., C.N.P., G.De., P.D., L.B., U.L., S.I.B., C.M.L., N.J.T., A.Ton., P.B.M., T.I.S., C.N.R., D.K.A., A.J.O., S.L.K., B.B., G.Ga., A.P.M., J.G.E., M.J.W., N.G.M., S.C.H., J.M.S., I.J.D., L.R.G., H.T., N.Pi., J.Ka., N.J.W., L.P., J.G.W., G.Gi., M.J.C., O.R., D.D.B., C.Gi., P.v.d.H., A.A.H., P.Kr., J.S., P.Kn., M.J., P.K.M., A.H., R.Sc., I.B.B., E.Va., D.M.B., D.B., K.L.M., M.B., C.M.v.D., D.K.S., A.Te., E.Z., A.Me., P.G., S.U., C.O., D.T., G.D.S., I.R., D.J.P., M.C., T.D.S., C.Hay., J.D., R.J.L., A.F.W., G.R.C., P.V., A.Sh., P.M.R., J.I.R., N.S., U.G., K.E.N., M.P., B.M.P., D.R.W., M.La., V.G., A.Ta., J.C.C., J.S.K., D.P.S., H.C., J.N.H., M.P., O.P. and J.F.W. designed individual studies. T.N., J.D.F., S.E., V.V., S.Tr., D.I.C., S.S.N., M.Ma., D.R., A.F., L.R.Y., E.H., C.Bo., J.R.P., S.C., U.B., G.M., T.Li., I.D., J.Z., J.P.B., E.S., S.Y., M.A.A., S.J.B., G.R.B., E.P.B., A.Ca., Y. Chan, S.J.C., Y.D.I.C., F.S.C., J.C., A.Co., L.Cu., G.Da., M.D., S.B.E., B.F., J.M.F., I.F., C.S.F., T.M.F., N.Fri., F.Ge., I.Gi., O.G., F.Gr., C.Gu., C.J.H., S.E.H., N.D.H., N.L.H., K.H., L.J.H., G.Ho., P.G.H., E.I., A.J., P.J., J.J., M.Ka., S.K., S.M.K., N.M.K., H.K.K., M.Ku., J.Ku., J.L., R.A.L., T.Le., D.C.L., L.Li., M.L.L., A.Lo., T.Lu., A.Lu., S.M., K.M., J.B.M., C.Mei., T.M., C.Men., F.D.M., L.M., G.W.M., R.H.M., R.N., M.N., M.S.N., G.T.O., A.O., S.P., W.R.P., J.S.P., I.Pa., K.P., N.Po., S.Ra., P.R., S.S.R., H.R., A.R., L.M.R., R.R., B.Sa., R.M.S., V.S., A.Sa., L.J.S., S.Se., P.S., B.H.S., D.R.W., M.La., V.G., A.Ta., J.C.C., J.S.K., D.P.S., H.C., J.N.H., M.P., O.P. and J.F.W. collected the data. S.Tr., D.I.C., M.C.C., C.Bo., U.B., I.D., M.A., F.W.A., S.J.B., D.J.B., E.B., E.P.B., A.Cc., S.J.C., J.C., I.F., T.M.F., C.Gu., C.J.H., T.B.H., N.D.H., M.J., E.I., J.J., P.Ko., M.Ku., L.J.L., R.A.L., L.Li., R.A.M., K.M., J.B.M., G.W.M., R.H.M., P.A.P., K.P., S.S.R., R.R., H.S., P.S., B.H.S., N.Sor., N.Sot., D.Va., J.B.W., C.S.Y., M.Me., N.J.S., D.C., D.A.M., R.S.C., P.F., G.P., S.F.G., H.H., L.F., G.De., P.D., L.B., U.L., S.I.B., C.M.L., A.Ton., P.B.M., C.N.R., D.K.A., A.J.O., S.L.K., B.B., G.Ga., A.P.M., M.J.W., N.G.M., S.C.H., J.M.S., I.J.D., L.R.G., J.Ka., N.J.W., L.P., J.G.W., G.Gi., M.J.C., O.R., D.D.B., C.Gi., P.v.d.H., A.A.H., P.Kr., J.S., P.Kn., M.J., P.K.M., A.H., R.Sc., I.B.B., E.Va., D.M.B., D.B., K.L.M., M.B., C.M.v.D., D.K.S., E.Z., A.Me., P.G., S.U., C.O., I.R., D.J.P., M.C., T.D.S., C.Hay., R.J.L., A.F.W., G.R.C., P.V., A.Sh., P.M.R., J.I.R., N.S., U.G., M.P., B.M.P., D.R.W., M.La., J.C.C., J.S.K., D.P.S., J.N.H., M.P., O.P. and J.F.W. contributed to funding. P.K.J., T.E., H.Ma., N.E., I.Ga., T.N., A.U.J., C.Sc., A.V.Sm., W.Zhan., Y.O., A.Stc., J.D.F., W. Zhao, T.M.B., M.P.C., N.Fra., S.E., V.V., S.Tr., X.G., D.I.C., J.R.O., T.C., S.S.N., Y. Chen, M.Ma., D.R., M.Ta., A.F., T.Kac., A.Bj., A.v.d.S., Y.W., A.K.G., L.R.Y., L.W., E.H., C.A.R., O.M., M.C.C., C.P., N.V., C.Ba., A.A.A., H.R.W., D.Vu., H.Me., J.R.P., S.S.Mi., M.C.B., S.S.Me., P.A.L., G.M., A.D., L.Y., L.F.B., D.Z., P.J.v.d.M., D.S., R.M., G.He., T.Kar., Z.W., T.Li., I.D., J.Z., W.M., L.La., S.W.v.L., J.P.B., A.R.W., A.Bo., T.S.A., L.M.H., E.S., S.Y., I.M.M., L.Ca., H.G.d.H., M.A., U.A., N.A., F.W.A., S.E.B., S.B., A.Ca., Y. Chan, C.C., G.Da., G.E., B.F., M.F.F., F.Ge., M.G., S.E.H., J.J.H., J.E.H., P.G.H., A.J., Y.K., S.K., R.A.L., B.L., M.Lo., S.J.Loo., Y.L., P.M., A.Ma., C.Men., F.D.M., E.M., M.E.M., A.Mo., A.O., I.Pa., F.P., I.Pr., L.M.R., B.Sa., R.M.S., R.Sa., H.S., W.R.S., C.Sa., C.Ma., B.Se., S.Sh., S.J.Lon., J.A.S., L.S., R.J.S., M.J.S., S.Ta., B.O.T., A.Tog., M.To., N.Ts., J.v.S., S.V., D.Vo., E.B.W., W.W., J.Y., G.Z., N.J.S., R.A.S., A.D.M., C.N.P., S.I.B., N.J.T., A.P.M., S.C.H., H.T., N.Pi., L.P., P.v.d.H., P.Kr., R.Sc., I.B.B., A.Te., C.O., M.C., J.D., J.I.R., N.S., K.E.N., A.Ta., J.C.C., J.S.K. and D.P.S. analysed the data. P.K.J., T.E., H.Ma., N.E., I.Ga., T.N., A.U.J., C.Sc., A.V.Sm., M.C.B. and D.P.S. performed beta-testing of scripts. P.K.J. and T.E. performed the meta-analysis. P.K.J., T.E., O.P. and J.F.W. wrote the manuscript. All authors approved the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Readers are welcome to comment on the online version of the paper. The authors declare competing financial interests: details are available in the online version of the paper. Correspondence and requests for materials should be addressed to J.F.W. (jim.wilson@ed.ac.uk).

METHODS

Outline. Our aim was to look for an association between a genetic effect (SROH) and 16 complex traits. Our approach followed best practice genome-wide association meta-analysis (GWAMA) protocols, where applicable, except we had only one genetic effect to test.

Cohorts were invited to join based on known previous participation in GWAMA and willingness to participate. 159 sub-cohorts were created from 102 population-based or case-control genetic studies, by separating different genotyping arrays, cases and controls or ethnic sub-groups to ensure each sub-cohort was homogeneous. Within each of the 159 sub-cohorts we measured the association between SROH and trait using the following model. Where a sub-cohort had been ascertained on the basis of a disease status associated with a particular trait, that sub-cohort was excluded from the corresponding trait analysis.

Phenotype was regressed on genetic effect and known relevant covariates within each cohort, under the model specified in equation (1). The estimated genetic effect of SROH was then meta-analysed using inverse variance meta-analysis.

$$Y = \mu + b_1 \text{ SROH} + b_2 \text{ age} + b_3 \text{ sex} + b_4 \text{ PC1} + b_5 \text{ PC2} + b_6 \text{ PC3} + e \quad (1)$$

Where Y is the vector of trait values, μ the intercept, b_1 the effect of SROH and b_2 – b_6 the effect of covariates. PC1–PC3, the post quality control within-cohort principal components of the cohort's relationship matrix and e the residual. Relationship matrices were determined genomically by each cohort using genome-wide array data. In addition, any other cohort-specific covariates known to be associated with the trait, including further principal components, and any trait-specific covariates and stratifications, such as medication and smoking status, were fitted as specified below. SROH was the sum of ROH called, with a length of at least 1.5 Mb using PLINK31.

As is routine in GWAMA, for family-based studies only, we also fitted an additional term to account for additive genetic values and relatedness, using grammar+ type residuals and full hierarchical mixed modelling using GenABEL³² and hglm³³, as specified in equation (2).

$$Y = \mu + b_1 \text{ SROH} + b_2 \text{ age} + b_3 \text{ sex} + b_4 \text{ PC1} + b_5 \text{ PC2} + b_6 \text{ PC3} + Za \quad (2)$$

Where a is the additive genetic value of each individual. $\text{Var}(a)$ is assumed to be proportionate to the genomic relationship matrix (GRM) (a pedigree relationship matrix was used in the Framingham Heart Study). Z is the identity matrix.

We then meta-analysed the regression coefficients (b_1) of traits on SROH for the 159 sub-cohorts.

Cohort recruitment. Data from 102 independent genetic epidemiology studies of adults were included. All subjects gave written informed consent and studies were approved by the relevant research ethics committees. Homogeneous sub-cohorts were created for analysis on the basis of ethnicity, genotyping array or other factors. Where a cohort had multiple ethnicities, sub-cohorts for each separate ethnicity were created and analysed separately. In all cases individuals of European, African, South or Central Asian, East Asian and Hispanic heritage individuals were separated. In some cases sub-categories, such as Ashkenazi Jews, were also distinguished. Ethnic outliers were excluded, as were the second of any monozygotic twins and pregnant subjects. Continental ancestry of cohorts participating in each trait study is presented in Extended Data Table 1. Cohort genotyping and summary information are shown in Supplementary Table 6, with age, sex, trait and homozygosity summary statistics given in Supplementary Tables 9, 10 and 11. For case-control and trait-extreme studies, patients or extreme-only sub-cohorts were analysed separately to controls. Where case status was associated with the trait under analysis the sub-cohort was excluded from that study (see below).

Subjects within a sub-cohort were genotyped using the same SNP array, or, where two very similar arrays were used (for example, Illumina OmniExpress and IlluminaOmni1), the intersection of SNPs on both arrays, provided the intersection exceeded 250,000 SNPs. Where a study used two different genotyping arrays, separate sub-cohorts were created for each array, and analysis was done separately. Paediatric cohorts were not included.

Genotyping. All subjects were genotyped using high-density genome-wide (>250,000 SNP) arrays, from Illumina, Affymetrix or Perlegen. Custom arrays were not included. Each study's usual array-specific genotype quality control standards for genome-wide association were used and are shown in Supplementary Table 6. Only autosomal data were analysed.

Phenotyping. We studied 16 quantitative traits which are widely available and represent different domains related to health, morbidity and mortality: height, body mass index (BMI), waist:hip ratio (WHR), diastolic and systolic blood pressure (DBP, SBP), fasting plasma glucose (FPG), fasting insulin (FI), haemoglobin A1c (HbA1c), total cholesterol, HDL and LDL cholesterol levels, triglycerides, forced expiratory volume in one second (FEV1), ratio of FEV1 to forced vital

capacity (FVC), general cognitive ability (g) and years of educational attainment. Phenotypic quality control was performed locally to assess the accuracy and distribution of phenotypes and covariates. Further covariates were included when the relevant genome-wide association study consortium also included them. The trait categories were anthropometry, blood pressure, glycaemic traits, classical lipids, lung function, cognitive function and educational attainment, following models in the GIANT³⁴, ICBP³⁵, MAGIC³⁶, CHARGE³⁷, Spirometa³⁸ and SSGAC³⁹ consortia. The model for FEV1 did not include height as a covariate. Effect sizes for FEV1 therefore include size effects that also underpin height. Studies assembled files containing study traits and the following covariates: sex, age, first three principal components of ancestry, lipid-lowering medication, ever-smoker status, anti-hypertensive medication, diabetes status and year of birth. Educational attainment was defined in accordance with the ICD 1997 classification (UNESCO), leading to seven categories of educational attainment that are internationally comparable³⁹. LDL values estimated using Friedewald's equation were accepted. Cohorts without fasting samples did not participate in the LDL-cholesterol, triglycerides, fasting insulin or fasting plasma glucose analyses. Cohorts with semi-fasting samples fitted a categorical or quantitative fasting time variable as a covariate. Subjects with less than 4 h fasting were not included.

Where subjects were ascertained, for example, on the basis of hypertension, that sub-cohort was excluded from analysis of traits associated with the disorder, for example blood pressure. The traits excluded from meta-analysis are as follows: ascertainment on type 2 diabetes, thus fasting insulin, HbA1c and fasting plasma glucose excluded; ascertainment on hypertension, thus blood pressures excluded; ascertainment on venous thrombosis or coronary artery disease, thus blood lipids excluded; ascertainment on obesity or the metabolic syndrome, thus blood lipids, body mass index, waist-hip ratio, fasting insulin and fasting plasma glucose excluded.

Somewhat unusually for a large consortium meta-analysis, the majority of the analysis after initial genotype and phenotype quality control was performed by a pipeline of standardised R and shell scripts, to ensure uniformity and reduce the risk of errors and ambiguities (available at <https://www.wiki.ed.ac.uk/display/ROHgen/Analysis+Plan+production+release+3.0>). The pipeline was used for all stages from this point onwards.

Calling runs of homozygosity. SNPs with more than 3% missingness across individuals or with a minor allele frequency less than 5% were removed. ROH were defined as runs of at least 50 consecutive homozygous SNPs spanning at least 1,500 kb, with less than a 1,000 kb gap between adjacent ROH and a density of SNP coverage within the ROH of no more than 50 kb/SNP, with one heterozygote and five no calls allowed per window, and were called using PLINK³¹, with the following settings: homozyg-window-snp 50; homozyg-snp 50; homozyg-kb 1500; homozyg-gap 1000; homozyg-density 50; homozyg-window-missing 5; homozyg-window-het 1. The same criteria were used by McQuillan *et al.*³, except SNP density has been relaxed to avoid regions of sparser coverage (still including 50 SNPs) being missed. The sum of runs of homozygosity (SROH) was then calculated. F_{ROH} was calculated as $\text{SROH}/(3 \times 10^9)$ reflecting the length of the autosomal genome. Copy number variants (CNV) are known to influence cognition⁴⁰; however, prior calling of CNV and ROH in one of our cohorts reduced the SROH by only 0.3%, making it implausible that deletions called as ROH influence our findings.

ROH called from different genotyping arrays. We show that SROH called with these parameters is relatively insensitive to the density and type of array used (Extended Data Fig. 7). We used 2.5 million SNPs available for 851 HapMap and 1000 Genomes Project⁴¹ samples from multiple continents to investigate the effect of array when using our ROH-calling parameters in PLINK. The data set included samples of African, European, admixed American, South and East Asian heritage. By subsampling SNPs from the 2.5 million we created array data for the commonly used Illumina CNV370 and OmniExpress beadchips and the Affymetrix6 array for each individual (see Supplementary Table 7 for details of the SNP numbers). The correlation in SROH using different arrays on the same individuals was 0.93–0.94 for all pairwise chip comparisons.

Trait association with SROH. The association between trait and SROH was calculated using a linear model in accordance with equation (1). Additional covariates were fitted for some analyses (shown below) or for some cohorts where analysts were aware of study specific effects (for example, study centre). For BMI, WHR, FEV1, FEV1/FVC and g , trait residuals were calculated for the model excluding SROH, these residuals were then rank-normalized and the effect of SROH on these rank-normalized residuals estimated. Triglycerides and fasting insulin were ln-transformed. Additional covariates were as follows: age² was included as a covariate for all traits apart from height and g . BMI was included as a covariate for WHR, SBP, DBP, FPG, FI and HbA1c. Year of birth was included as a covariate for educational attainment and ever-smoking for FEV1 and FEV1/FVC. Where a subject was known to be taking lipid-lowering medication, total

cholesterol was adjusted by dividing by 0.8. Similarly, where a subject was known to be taking anti-hypertensive medication, SBP and DBP measurements were increased by 15 and 10 mm Hg, respectively.

Where the cohort was known to have significant kinship, genetic relatedness was also fitted, using the mixed model, in accordance with equation (2). The polygenic model was fitted in GenABEL using the fixed covariates and the genomic relationship matrix³². GRAMMAR+ (GR+) (ref. 42) residuals were then fitted to SROH as well as the full mixed model being fitted simultaneously, using GenABEL's hierarchical generalized linear model (HGLM) function³³. Populations with kinship thus potentially had three estimates of $\beta_{F_{ROH}}$: using fixed effects only, and using the mixed model approaches, (GR+ and HGLM) for SROH.

To investigate potential confounding, where available, educational attainment was added as an ordinal covariate and all models rerun, giving revised estimates of $\beta_{F_{ROH}}$. This is potentially an over adjustment for g due to the phenotypic and genetic correlations with educational attainment⁴³. However it must be recognized that educational attainment does not capture all potential environmental confounding.

Cohort phenotypic means and standard deviations were checked visually for inter-cohort consistency, with apparent outliers then being corrected (for example, due to units or incorrectly specified missing values), explained (for example, due to different population characteristics) or excluded. Individual sub-cohort trait means and standard deviations are tabulated in Supplementary Table 9 and age and gender information is in Supplementary Table 10.

Meta-analysis. As is routine in genome-wide association meta-analyses, analysis was performed within homogeneous sub-populations and only meta-analysis of the estimated (within-population) effect sizes was used to combine results between populations, avoiding any confounding effects of inter-population differences in trait or genetic effect distributions. Inverse-variance meta-analysis of all sub-cohorts' effect estimates was performed using Rmeta, on a fixed-effect basis (Supplementary Table 5 compares random effects meta-analysis). In the principal analyses, for cohorts with relatedness, HGLM estimates of $\beta_{F_{ROH}}$ were preferred; however, where HGLM had failed to converge, results using GRAMMAR+ were included. These results were combined with those for unrelated cohorts on a fixed-model-only basis. Result outliers were defined as individual cohort by trait results, which failed the hypothesis, cohort ($\beta_{F_{ROH}}$) = pre-quality-control meta-analysis ($\beta_{F_{ROH}}$), with a t -test statistic >3 . Analyses were performed with and without outliers for $\beta_{F_{ROH}}$ in phenotypic units and in intra-sex phenotypic standard deviations (Supplementary Table 8). The principal results we present are for F_{ROH}

with outliers included for the hypothesis tests (which turns out to be more conservative), but with outliers excluded when estimating $\beta_{F_{ROH}}$ (ref. 44). Meta-analysis was performed using inverse variance meta-analysis in the R package Rmeta, with $\beta_{F_{ROH}}$ taken as a fixed effect and alternatively as a random effect. The principal results are on a fixed-effects basis, with Supplementary Table 5 showing comparison with the random-effects analysis.

Meta-analyses were re-run for various subsets, according to geographic and demographic features of the cohorts. Cohorts were divided into more homozygous and less homozygous strata with the boundary being set so each within-stratum meta-analysis had equal statistical power.

Data reporting. Randomization and blind allocation were not applicable to this study.

31. Purcell, S. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
32. Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–1296 (2007).
33. Ronnegard, L., Shen, X. & Alam, M. hglm: a package for fitting hierarchical generalized linear models. *R Journal* **2**, 20–28 (2010).
34. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
35. Ehret, G. B. *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
36. Scott, R. A. *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nature Genet.* **44**, 991–1005 (2012).
37. Willer, C. J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nature Genet.* **45**, 1274–1283 (2013).
38. Soler Artigas, M. *et al.* Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nature Genet.* **43**, 1082–1090 (2011).
39. Rietveld, C. A. *et al.* GWAS of 126,559 individuals identified genetic variants associated with educational attainment. *Science* **340**, 1467–1471 (2013).
40. Stefansson, H. *et al.* CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature* **505**, 361–366 (2014).
41. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
42. Aulchenko, Y. S., de Koning, D. J. & Haley, C. Genomewide rapid association using mixed model and regression: a fast and simple method for genome-wide pedigree-based quantitative trait loci association analysis. *Genetics* **177**, 577–585 (2007).
43. Marioni, R. E. *et al.* Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence* **44**, 26–32 (2014).
44. Hedges, L. V. & Olkin, I. *Statistical Methods for Meta-Analysis* (Academic Press, New York, 1985).

- Peter K. Joshi^{1*}, Tonu Esko^{2,3,4,5*}, Hannele Mattsson^{6,7}, Niina Eklund⁶, Ilaria Gandin⁸, Teresa Nutile⁹, Anne U. Jackson¹⁰, Claudia Schurmann^{11,12}, Albert V. Smith^{13,14}, Weihua Zhang^{15,16}, Yukinori Okada^{17,18}, Alena Stančáková¹⁹, Jessica D. Faul²⁰, Wei Zhao²¹, Traci M. Bartz²², Maria Pina Concas²³, Nora Franceschini²⁴, Stefan Enroth²⁵, Veronique Vitart²⁶, Stella Markovitz²⁷, Xiuqing Guo^{28,29}, Daniel I. Chasman³⁰, Jeffrey R. O'Connell³¹, Tanguy Corre^{32,33}, Suraj S. Nongmaithem³⁴, Yuning Chen³⁵, Massimo Mangino^{36,37}, Daniela Ruggiero³⁸, Michela Traglia³⁹, Alike-Eleni Farmaki³⁹, Tim Kacprowski⁴⁰, Andrew Bjorndal⁴¹, Ashley van der Spek⁴², Ying Wu⁴³, Anil K. Giri⁴⁴, Lisa R. Yanek⁴⁵, Lihua Wang⁴⁶, Edith Hofer^{47,48}, Cornelius A. Rietveld⁴⁹, Olga McLeod⁵⁰, Marilyn C. Cornelis^{51,52}, Cristian Pattaro⁵³, Niek Verweij⁵⁴, Clemens Baumbach^{55,56,57}, Abdel Abdellaoui⁵⁸, Helen R. Warren^{59,60}, Dragana Vuckovic⁸, Hao Mei⁶¹, Claude Bouchard⁶², John R. B. Perry⁶³, Stefania Cappellani⁶⁴, Saira S. Mirza⁴², Miles C. Benton⁶⁵, Ulrich Broeckel⁶⁶, Sarah E. Medland⁶⁷, Penelope A. Lind⁶⁷, Giovanni Malerba⁶⁸, Alexander Drong⁶⁹, Loic Yengo⁷⁰, Lawrence F. Bielak⁷¹, Degui Zhi⁷¹, Peter J. van der Most⁷², Daniel Shriver⁷³, Reedik Mägi⁷³, Gibran Hemani⁷⁴, Tugce Karaderi⁶⁹, Zhaoxing Wang^{75,76}, Tian Liu^{77,78}, Jia Demuth^{79,80}, Jing Hua Zhao⁶³, Weihua Meng⁸¹, Lazaros Latiotakis⁸², Sander W. van der Laan⁸³, Jonathan P. Bradfield⁸⁴, Andrew R. Wood⁸⁵, Amelie Bonnefond⁷⁰, Tarunveer S. Ahluwalia^{86,88,232}, Leanne M. Hall⁸⁹, Erika Salvi⁹⁰, Seyhan Yazar⁹¹, Lisbeth Carstensen⁹², Hugoline G. de Haan⁹³, Mark Abney⁹⁴, Uzma Afzal^{15,16}, Matthew A. Allison⁹⁵, Najaf Amin⁴², Folkert W. Asselbergs^{96,97,98}, Stephan J. L. Bakker⁹⁹, R. Graham Barr¹⁰⁰, Sebastian E. Baumeister¹⁰¹, Daniel J. Benjamin^{102,103}, Sven Bergmann^{32,33}, Eric Boerwinkle¹⁰⁴, Erwin P. Bottinger¹¹, Archie Campbell¹⁰⁵, Aravinda Chakravarti¹⁰⁶, Yingleong Chan^{34,5}, Stephen J. Chanock⁷⁵, Constance Chen¹⁰⁷, Y. -D. Ida Chen^{28,29}, Francis S. Collins¹⁰⁸, John Connell¹⁰⁹, Adolfo Correa⁶¹, L. Adrienne Cupples^{35,110}, George Davey Smith⁷⁴, Gail Davies^{111,112}, Marcus Dörn¹¹³, Georg Ehret^{106,114}, Stephen B. Ellis¹¹, Bjarke Feenstra⁹², Mary F. Feitosa⁴⁶, Ian Ford¹¹⁵, Caroline S. Fox^{110,116}, Timothy M. Frayling⁸⁵, Nele Friedrich¹¹⁷, Frank Geller⁹², Generation Scotland¹⁰⁵, Irina Gillham-Naseny³⁶, Omri Gottesman¹¹, Misa Graff²⁴, Francine Grodstein⁵², Charles Gu¹¹⁸, Chris Haley^{26,119}, Christopher J. Hammond³⁶, Sarah E. Harris^{105,112}, Tamara B. Harris¹²⁰, Nicholas D. Hastie²⁶, Nancy L. Heard-Costa^{110,121}, Kauko Heikkilä¹²², Lynne J. Hocking¹²³, Georg Homuth⁴⁰, Jouke-Jan Hottenga⁵⁸, Jinyan Huang¹²⁴, Jennifer E. Huffman²⁶, Pirro G. Hysi³⁶, M. Arfan Ikram^{42,125}, Erik Ingelsson^{69,126}, Anni Joensuu^{6,7}, Åsa Johansson^{25,127}, Pekka Jousilahti¹²⁸, J. Wouter Jukema¹²⁹, Mika Kähönen¹³⁰, Yoichi Kamatani¹⁸, Stavroula Kononi⁸², Shona M. Kerr²⁶, Nazir M. Khan⁴⁴, Philipp Koellinger⁴⁹, Heikki A. Koistinen^{131,132,133}, Manraj K. Kooner¹⁶, Michiaki Kubo¹³⁴, Johanna Kuusisto¹³⁵, Jari Lahti^{136,137}, Lenore J. Launer¹²⁰, Rodney A. Lea⁶⁵, Benjamin Lehne¹⁵, Terho Lehtimäki¹³⁸, David C.M. Liewald¹¹², Lars Lind¹³⁹, Marie Loh^{15,233}, Marja-Liisa Lokki¹⁴⁰, Stephanie J. London¹⁴¹, Stephanie J. Loomis¹⁴², Anu Loukola¹²², Yingchang Lu^{11,12}, Thomas Lumley¹⁴³, Annamari Lundqvist¹⁴⁴, Satu Männistö¹²⁸, Pedro Marques-Vidal¹⁴⁵, Corrado Masciullo³⁸, Angela Matchan¹⁴⁶, Rasika A. Mathias^{45,147}, Koichi Matsuda¹⁴⁸, James B. Meigs¹⁴⁹, Christa Meisinger⁵⁶, Thomas Meitinger^{150,151}, Cristina Menni³⁶, Frank D. Mentch⁸⁴, Evelin Mihailov², Lili Milani², May E. Montasser³¹, Grant W. Montgomery¹⁵², Alanna Morrison¹⁰⁴, Richard H. Myers¹⁵³, Rajiv Nadukuru¹¹, Pau Navarro²⁶, Mari Nelis², Markku S. Nieminen¹⁵⁴, Ilja M. Nolte⁷², George T. O'Connor^{101,155}, Adesola Ogundimu¹⁵⁶, Sandosh Padmanabhan¹⁵⁷, Walter R. Palmas¹⁰⁰, James S. Pankow¹⁵⁸, Inga Patarcic¹⁵⁹, Francesca Pavan⁵³, Patricia A. Peyser²¹, Kirsi Pietiläinen^{17,132,160}, Neil Poulter¹⁶¹, Inga Prokopenko¹⁶², Sarju Ralhan¹⁶³, Paul Redmond¹¹¹, Stephen S. Rich¹⁶⁴, Harri Rissanen¹⁴⁴, Antonietta Robino⁶⁴, Lynda M. Rose³⁰, Richard Rose¹⁶⁵, Cinzia Sala³⁸, Babatunde Salako¹⁵⁶, Veikko Salomaa¹²⁸, Antti-Pekka Sarin⁶⁷, Richa Saxena⁴¹, Helena Schmidt¹⁶⁶, Laura J. Scott¹⁰, William R. Scott^{15,16}, Bengt Sennblad^{50,167}, Sudha Seshadri^{110,121}, Peter Sever¹⁶¹, Smeeta Shrestha³⁴, Blair H. Smith¹⁶⁸, Jennifer A. Smith²¹, Nicole Soranzo¹⁴⁶, Nona Sotoodehnia¹⁶⁹, Lorraine Southam^{69,146}, Alice V. Stanton¹⁷⁰, Maria G. Stathopoulou¹⁷¹, Konstantin Strauch^{57,172}, Rona J. Strawbridge⁵⁰, Matthew J. Suderman⁷⁴, Nikhil Tandon¹⁷³, Sian-Tsun Tang¹⁷⁴, Kent D. Taylor^{28,29}, Bamidele O. Tayo¹⁷⁵, Anna Maria Töglhofer¹⁶⁶, Maciej Tomaszewski^{89,176}, Natalia Tsernikova¹⁷⁷, Jaakko Tuomilehto^{131,178,179}, Andre G. Uitterlinden^{42,180}, Dhananjay Vaidya^{45,181}, Astrid van Hylckama Vlieg⁹³, Jessica van Setten⁸³, Tuula Vasankari¹⁸², Sailaja Vedantam^{34,5}, Eftymia Vlachopoulou¹⁴⁰, Diego Vozza⁶⁴, Eero Vuoksimaa¹²², Melanie Waldenberger^{55,56}, Erin B. Ware²¹, William Wentworth-Shields⁹⁴, John B. Whitfield¹⁸³, Sarah Wild³⁴, Gonke Willemsen⁵⁸, Chittaranjan S. Yajnik¹⁸⁴, Jie Yao²⁸, Gianluigi Zaza¹⁸⁵, Xiaofeng Zhu¹⁸⁶, The BioBank Japan Project¹⁸, Rany M. Salem^{34,5}, Mads Melbye^{92,187}, Hans Bisgaard⁸⁶, Nilesh J. Samani^{89,176}, Daniele Cusi⁹⁰, David A. Mackey⁹¹, Richard S. Cooper¹⁷⁵, Philippe Froguel^{70,162}, Gerard Pasterkamp⁸³, Struan F.A. Grant^{84,188}, Hakon Hakonarson^{84,188}, Luigi Ferrucci¹⁸⁹, Robert A. Scott⁶³, Andrew D. Morris¹⁹⁰, Colin N. A. Palmer¹⁹¹, George Dedoussis³⁹, Panos Deloukas^{82,192}, Lars Bertram^{78,193,235}, Ulman Lindenberg⁷⁷, Sonja I. Berndt⁷⁵, Cecilia M. Lindgren^{4,69}, Nicholas J. Timpson⁷⁴, Anke Tönjes¹⁹⁴, Patricia B. Munroe^{59,60}, Thorikild I. A. Sørensen^{74,88,195}, Charles N. Rotimi⁷³, Donna K. Arnett¹⁹⁶, Albertine J. Oldehinkel¹⁹⁷, Sharon L. R. Kardia²¹, Beverley Balkau¹⁹⁸, Giovanni Gambaro¹⁹⁹, Andrew P. Morris^{2,69,200}, Johan G. Eriksson^{128,201,202,203,204}, Margie J. Wright²⁰⁵, Nicholas G. Martin¹⁸³, Steven C. Hunt²⁰⁶, John M. Starr^{112,207}, Ian J. Deary^{111,112}, Lyn R. Griffiths⁶⁵, Henning Tiemeier^{42,208}, Nicola Pirastu^{8,64}, Jaakko Kaprio^{7,122,209}, Nicholas J. Wareham⁶³, Louis Pérusse²¹⁰, James G. Wilson²¹¹, Giorgio Grotto⁷, Mark J. Caulfield^{59,60}, Olli Raitakari^{212,213}, Dorret I. Boomsma⁵⁸, Christian Gieger^{55,56,57}, Pim van der Harst^{54,97,214}, Andrew A. Hicks⁵³, Peter Kraft¹⁰⁷, Juha Sinisalo¹⁵⁴, Paul Knekt¹⁴⁴, Magnus Johannesson²¹⁵, Patrik K. E. Magnusson²¹⁶, Anders Hamsten⁵⁰, Reinhold Schmidt⁴⁷, Ingrid B. Borecki²¹⁷, Erkki Vartiainen¹²⁸, Diane M. Becker^{45,218}, Dwaipayan Bhardwaj^{44,236}, Karen L. Mohlke⁴³, Michael Boehnke¹⁰, Cornelia M. van Duijn⁴², Dharamvir K. Sanghera^{219,220}, Alexander Teumer¹⁰¹, Eleftheria Zeggini¹⁴⁶, Andres Metspalu^{2,177}, Paolo Gasparini^{64,221}, Sheila Ulivi⁶⁴, Carole Ober⁹⁴, Daniela Toniolo³⁸, Igor Rudan²¹, David J. Porteous^{105,112}, Marina Ciullo⁷, Tim D. Spector³⁶, Caroline Hayward²⁶, José Dupuis^{35,110}, Ruth J. F. Loos^{11,12,222}, Alan F. Wright²⁶, Giriraj R. Chandak^{34,223}, Peter Vollenweider¹⁴⁵, Alan R. Shuldiner^{31,224,225}, Paul M.

Ridker³⁰, Jerome I. Rotter^{28,29}, Naveed Sattar²²⁶, Ulf Gyllenstein²⁵, Kari E. North^{24,227}, Mario Pirastu²³, Bruce M. Psaty^{228,229}, David R. Wei²⁰, Markku Laakso¹³⁵, Vilhelm Gudnason^{131,14}, Atsushi Takahashi¹⁸, John C. Chambers^{15,16,230}, Jaspal S. Kooner^{16,174,230}, David P. Strachan²³¹, Harry Campbell¹, Joel N. Hirschhorn^{3,4,5}, Markus Perola^{2,6}, Ozren Polasek^{1,159*}, & James F. Wilson^{1,26*}

- ¹Centre for Global Health Research, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, UK. ²Estonian Genome Center, University of Tartu, Riia 23b, 51010, Tartu, Estonia. ³Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Cambridge, 02141 Massachusetts, USA. ⁴Program in Medical and Population Genetics, Broad Institute, Cambridge Center 7, Cambridge, Massachusetts 02242, USA. ⁵Department of Genetics, Harvard Medical School, 25 Shattuck St, Boston, Massachusetts 02115, USA. ⁶Unit of Public Health Genomics, National Institute for Health and Welfare, P.O. Box 104, Helsinki, FI-00251, Finland. ⁷Institute for Molecular Medicine Finland (FIMM), University of Helsinki, P.O. Box 20, Helsinki, FI-00014, Finland. ⁸Department of Medical Sciences, University of Trieste, Strada di Fiume 447 - Osp. di Cattinara, 34149 Trieste, Italy. ⁹Institute of Genetics and Biophysics "A. Buzzati-Traverso" CNR, via Pietro Castellino, 111, 80131 Naples, Italy. ¹⁰Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA. ¹¹The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, New York 10029, USA. ¹²The Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, New York 10029, USA. ¹³Icelandic Heart Association, Holtasmari 1, 201, Kopavogur, Iceland. ¹⁴Faculty of Medicine, University of Iceland, Reykjavik, 101, Iceland. ¹⁵Department of Epidemiology and Biostatistics, Imperial College London, Norfolk Place, London W2 1PG, UK. ¹⁶Department of Cardiology, Ealing Hospital NHS Trust, Uxbridge Road, Southall, Middlesex UB1 3HW, UK. ¹⁷Department of Human Genetics and Disease Diversity, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo, 113-8510, Japan. ¹⁸Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan. ¹⁹Department of Medicine, University of Eastern Finland, 70210 Kuopio, Finland. ²⁰Institute for Social Research, University of Michigan, 426 Thompson Street, Ann Arbor, Michigan 48104, USA. ²¹Department of Epidemiology, University of Michigan, 1415 Washington Heights, Ann Arbor, Michigan 48109, USA. ²²Cardiovascular Health Research Unit, Departments of Biostatistics and Medicine, University of Washington, 1730 Minor Ave, Suite 1360, Seattle, Washington 98101, USA. ²³Institute of Population Genetics, National Research Council, Trav. La Crucca n. 3 - Reg. Balduina, 07100 Sassari, Italy. ²⁴Epidemiology, University of North Carolina, 137 E. Franklin St, Suite 306, Chapel Hill, North Carolina 27599, USA. ²⁵Department of Immunology, Genetics, and Pathology, Biomedical Center, SciLifeLab Uppsala, Uppsala University, SE-75108 Uppsala, Sweden. ²⁶MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Crewe Road, EH4 2XU Edinburgh, UK. ²⁷Department of Gerontology and Geriatrics, Leiden University Medical Center, PO Box 9600, Leiden, 2300 RC, The Netherlands. ²⁸Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute, 1124 W. Carson Street, Torrance, California 90502, USA. ²⁹Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, California 90502, USA. ³⁰Division of Preventive Medicine, Brigham and Women's Hospital, 900 Commonwealth Avenue, East, Harvard Medical School, Boston, Boston, Massachusetts 02215, USA. ³¹Division of Endocrinology, Diabetes, and Nutrition and Program for Personalised and Genomic Medicine, Department of Medicine, University of Maryland School of Medicine, 685 Baltimore St. MSTF, Baltimore, Maryland 21201, USA. ³²Department of Medical Genetics, University of Lausanne, Rue du Bugnon 27, Lausanne, 1005, Switzerland. ³³Swiss Institute of Bioinformatics, Quartier Sorge - batiment génopode, Lausanne, 1015, Switzerland. ³⁴Genomic Research on Complex Diseases (GRC) Group, CSIR-Centre for Cellular and Molecular Biology, Habshiguda, Uppal Road, Hyderabad, 500007, India. ³⁵Department of Biostatistics, Boston University School of Public Health, 801 Massachusetts Avenue, Boston, Massachusetts 02118, USA. ³⁶Department of Twin Research & Genetic Epidemiology, King's College London, South Wing, Block D, 3rd Floor, Westminster Bridge Road, London SE1 7EH, UK. ³⁷NIHR Biomedical Research Centre, Guy's and St. Thomas' Foundation Trust, Westminster Bridge Road, London SE1 7EH, UK. ³⁸Division of Genetics and Cell Biology, San Raffaele Scientific Institute, Via Olgettina 58, 20132 Milano, Italy. ³⁹Department of Nutrition and Dietetics, Harokopio University of Athens, 70, El. Venizelou Ave, Athens 17671, Greece. ⁴⁰Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Friedrich-Ludwig-Jahn-Str. 15A, Greifswald 17475, Germany. ⁴¹Center for Human Genetic Research, 55 Fruit Street, Massachusetts General Hospital, Massachusetts 02114, USA. ⁴²Department of Epidemiology, Erasmus Medical Center, PO Box 2040, Rotterdam, 3000 CA, The Netherlands. ⁴³Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599, USA. ⁴⁴Genomics and Molecular Medicine, CSIR-Institute of Genomics & Integrative Biology, Mathura Road, New Delhi, 110025, India. ⁴⁵The GeneSTAR Research Program, Division of General Internal Medicine, Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21287, USA. ⁴⁶Department of Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, Saint Louis, Missouri 63108, USA. ⁴⁷Department of Neurology, Clinical Division of Neurogeriatrics, Medical University Graz, Auenbruggerplatz 22, Graz, A-8036, Austria. ⁴⁸Institute for Medical Informatics, Statistics and Documentation, Medical University Graz, Auenbruggerplatz 22, Graz, A-8036, Austria. ⁴⁹Erasmus School of Economics, Erasmus University Rotterdam, Burgemeester Oudlaan 50, Rotterdam, 3000 DR, The Netherlands. ⁵⁰Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, CMM L8:03, Karolinska University Hospital, Solna, Stockholm, 171 76, Sweden. ⁵¹Channing Division of Network Medicine, Brigham & Women's Hospital, 181 Longwood, Boston, Massachusetts 02115, USA. ⁵²Nutrition, Harvard School of Public Health, 401 Park Drive, Boston, Massachusetts 02215, USA. ⁵³Center for Biomedicine, European Academy Bozen/Bolzano (EURAC), 39100 Bolzano,

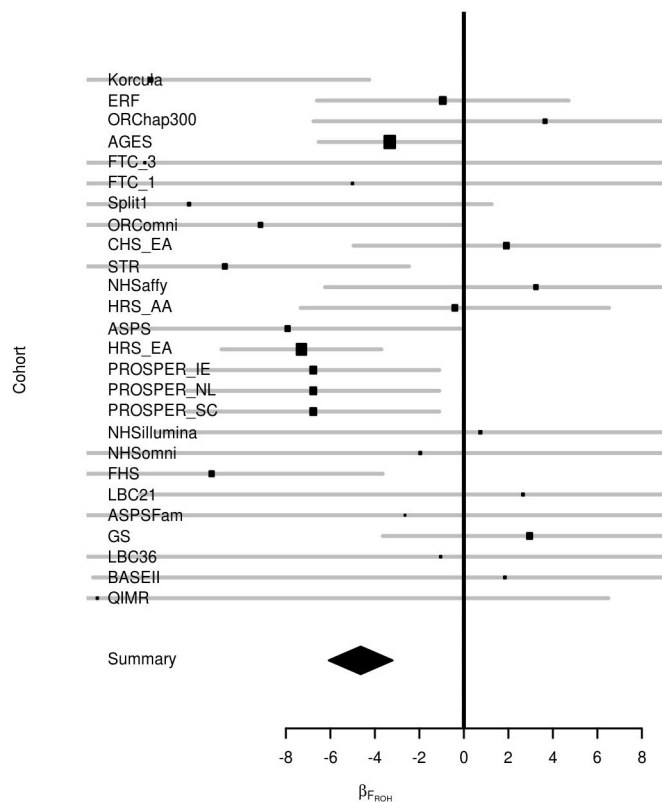
Italy (affiliated Institute of the University of Lübeck, D-23562 Lübeck, Germany).

⁵⁴University of Groningen, University Medical Center Groningen, Department of Cardiology, Hanzplein 1, Groningen, 9700 RB, The Netherlands. ⁵⁵Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstr. 1, Neuherberg 85764, Germany. ⁵⁶Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstr. 1, Neuherberg 85764, Germany. ⁵⁷Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstr. 1, Neuherberg 85764, Germany. ⁵⁸Department of Biological Psychology, VU University Amsterdam, Van der Boechorststraat 1, Amsterdam, 1081 BT, The Netherlands. ⁵⁹Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK. ⁶⁰NIHR Barts Cardiovascular Biomedical Research Unit, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK. ⁶¹Department of Medicine, University of Mississippi Medical Center, 2500 N. State St., Jackson, Mississippi 39216, USA. ⁶²Pennington Biomedical Research Center, 6400 Perkins Rd, Baton Rouge, Louisiana 70808, USA. ⁶³MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK. ⁶⁴Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", via dell'Istria 65, 34137 Trieste, Italy. ⁶⁵Institute of Health and Biomedical Innovation, Queensland University of Technology, 60 Musk Avenue, Kelvin Grove, GPO Box 2434, Brisbane Queensland 4001, Australia. ⁶⁶Department of Pediatrics, Medical College of Wisconsin, 8701 Watertown Plank Rd, Milwaukee, Wisconsin 53226, USA. ⁶⁷Quantitative Genetics, QIMR Berghofer Medical Research Institute, 300 Herston Rd, Herston, Brisbane Queensland 4006, Australia. ⁶⁸Dipartimento di Scienze della Vita e della Riproduzione, University of Verona, Strada Le Grazie 15, 37134 Verona, Italy. ⁶⁹Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK. ⁷⁰CNRS UMR 8199, European Genomic Institute for Diabetes (EGID), Lille 2 University, 1 Rue du Professeur Calmette, 59000 Lille, France. ⁷¹Department of Biostatistics, University of Alabama at Birmingham, 1665 University Blvd, Birmingham, Alabama 35294, USA. ⁷²Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, P.O. box 30.001, 9700 RB, Groningen, The Netherlands. ⁷³Center for Research on Genomics and Global Health, National Human Genome Research Institute, Building 12A/Room 4047, 12 South Dr., Bethesda, Maryland 20892, USA. ⁷⁴MRC Integrative Epidemiology Unit, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK. ⁷⁵Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, 9609 Medical Center Drive, Rockville, Maryland 20850, USA. ⁷⁶Cancer Genomics Research Laboratory, National Cancer Institute, SAIC-Frederick, Inc., Frederick National Laboratory for Cancer Research, Frederick, Maryland 21702, USA. ⁷⁷Center for Lifespan Psychology, Max Planck Institute for Human Development, Lentzeallee 94, Berlin 14195, Germany. ⁷⁸Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestr. 72, Berlin, 14195 Germany. ⁷⁹Charité Research Group on Geriatrics, Charité – Universitätsmedizin Berlin, Reinickendorferstr. 61, 13347 Berlin, Germany. ⁸⁰Institute of Medical and Human Genetics, Charité – Universitätsmedizin Berlin, Augustenburger Platz 1, Berlin 13353, Germany. ⁸¹Division of Population Health Sciences, Medical Research Institute, University of Dundee, Ninewells Hospital and School of Medicine, Dundee DD2 4BF, UK. ⁸²William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, Charterhouse Square, London, EC1M 6BQ, UK. ⁸³Experimental Cardiology, Division Heart and Lungs, University Medical Center Utrecht, Heidelberglaan 100, Utrecht, 3584 CX, The Netherlands. ⁸⁴Center for Applied Genomics, Children's Hospital of Philadelphia, 3615 Civic Center Boulevard, Philadelphia, Pennsylvania 19104, USA. ⁸⁵Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Royal Devon and Exeter Hospital, Barrack Road, Exeter EX2 5DW, UK. ⁸⁶COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Ledreborg Allé 34, DK-2820 Copenhagen, Denmark. ⁸⁸Nordisk Centre for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Universitetsparken 1, Copenhagen, 2100, Denmark. ⁸⁹Department of Cardiovascular Sciences, University of Leicester, BHF Cardiovascular Research Centre, Glenfield Hospital, Groby Road, Leicester LE3 9QP, UK. ⁹⁰Department of Health Sciences, University of Milan, via A. di Rudini 8, 20142 Milan, Italy. ⁹¹Centre for Ophthalmology and Visual Science, University of Western Australia, Lions Eye Institute, 2 Verdun Street, Perth, Western Australia 6009, Australia. ⁹²Department of Epidemiology Research, Statens Serum Institut, Artillerivej 5, Copenhagen, 2300, Denmark. ⁹³Clinical Epidemiology, Leiden University Medical Center, PO Box 9600, Leiden, 2300 RC, The Netherlands. ⁹⁴Department of Human Genetics, University of Chicago, 920 E. 58th Street, Chicago, Illinois 60637, USA. ⁹⁵Department of Family and Preventive Medicine, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA. ⁹⁶Department of Cardiology, Division Heart and Lungs, University Medical Center Utrecht, Heidelberglaan 100, Utrecht, 3584 CX, The Netherlands. ⁹⁷Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Catharijnesingel 52, Utrecht, 3501 DG, The Netherlands. ⁹⁸Institute of Cardiovascular Science, faculty of Population Health Sciences, University College London, Gower Street, London WC1E 6BT, UK. ⁹⁹University of Groningen, University Medical Center Groningen, Department of Internal Medicine, Hanzplein 1, Groningen, 9700 RB, The Netherlands. ¹⁰⁰Department of Medicine, Columbia University, 622 W. 168th Street, New York, New York 10032, USA. ¹⁰¹Institute for Community Medicine, University Medicine Greifswald, W.-Rathenau-Str. 48, Greifswald 17475, Germany. ¹⁰²Department of Economics, Cornell University, 480 Uris Hall, Ithaca, New York 14853, USA. ¹⁰³Department of Economics and Center for Economic and Social Research, University of Southern California, 314C Dauterive Hall, 635 Downey Way, Los Angeles, California 90089, USA. ¹⁰⁴Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, 1200 Pressler Street, Suite 453E, Houston, Texas 77030, USA. ¹⁰⁵Centre for Genomic and Experimental Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK. ¹⁰⁶McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ¹⁰⁷Program in Genetic Epidemiology and

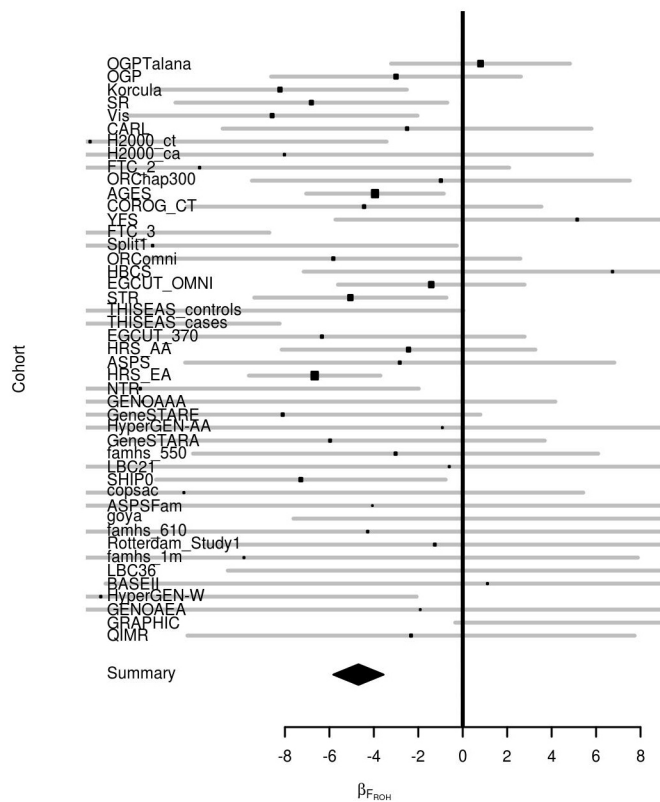
Statistical Genetics, Harvard School of Public Health, 665 Huntington Ave, Boston, Massachusetts 02115, USA. ¹⁰⁸Genome Technology Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892, USA. ¹⁰⁹College of Medicine, Dentistry and Nursing, Ninewells Hospital and Medical School, College Office, Level 10, Dundee DD1 9SY, UK. ¹¹⁰National Heart, Lung, and Blood Institute's Framingham Heart Study, 73 Mt. Wayte Ave, Framingham, Massachusetts 01702, USA. ¹¹¹Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK. ¹¹²Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK. ¹¹³Department of Internal Medicine B, University Medicine Greifswald, Ferdinand-Sauerbruch-Str. NK, Greifswald 17475, Germany. ¹¹⁴Cardiology, Geneva University Hospitals, Rue Gabrielle-Perret-Gentil, 4, Genève 14, 1211, Switzerland. ¹¹⁵Robertson Centre, University of Glasgow, Boyd Orr Building, Glasgow G12 8QQ, Scotland. ¹¹⁶Division of Endocrinology, Brigham and Women's Hospital and Harvard Medical School, 75 Francis St, Boston, Massachusetts 02115, USA. ¹¹⁷Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Ferdinand-Sauerbruch-Str. NK, 17475 Greifswald, Germany. ¹¹⁸Division of Biostatistics, Washington University, 660 S Euclid, St Louis, Missouri 63110, USA. ¹¹⁹Roslin Institute, University of Edinburgh, Easter Bush, Midlothian, Edinburgh EH25 9RG, UK. ¹²⁰National Institutes on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA. ¹²¹Department of Neurology, Boston University School of Medicine, 72 E Concord St, Boston, Massachusetts 02118, USA. ¹²²Department of Public Health, University of Helsinki, Hietalahti, P.O.Box 41, Mannerheimintie 172, Helsinki, FI-00014, Finland. ¹²³Musculoskeletal Research Programme, Division of Applied Medicine, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, UK. ¹²⁴State Key Laboratory of Medical Genomics, Shanghai Institute of Hematology, Rui Jin Hospital Affiliated with Shanghai Jiao Tong University School of Medicine, 197 Rui Jin Er Road, Shanghai, 200025, China. ¹²⁵Department of Radiology, Erasmus Medical Center, PO Box 2040, Rotterdam, 3000 CA, The Netherlands. ¹²⁶Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, SE-17121, Sweden. ¹²⁷Uppsala Clinical Research Center, Uppsala University, Uppsala, SE-75237, Sweden. ¹²⁸Department of Chronic Disease Prevention, National Institute for Health and Welfare, P.O. Box 30, Helsinki, FI-00271, Finland. ¹²⁹Department of Cardiology C5-P, Leiden University Medical Center, PO Box 9600, Leiden, 2300 RC, The Netherlands. ¹³⁰Department of Clinical Physiology, University of Tampere and Tampere University Hospital, P.O. Box 2000, Tampere, FI-33521, Finland. ¹³¹Diabetes Prevention Unit, National Institute for Health and Welfare, P.O. Box 30, FI-00271 Helsinki, Finland. ¹³²Department of Medicine, Division of Endocrinology, Helsinki University Central Hospital, P.O.Box 340, Haartmaninkatu 4, Helsinki, FI-00029, Finland. ¹³³Minerva Foundation Institute for Medical Research, Biomedicum 2U, Tukholmankatu 8, Helsinki, FI-00290, Finland. ¹³⁴Laboratory for Genotyping Development RCMIS, 1-7-22 Suehirocho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan. ¹³⁵Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, FI-70210, Finland. ¹³⁶Institute of Behavioural Sciences, University of Helsinki, P.O. Box 9, University of Helsinki, Helsinki, FI-00014, Finland. ¹³⁷Folkhälsan Research Centre, PB 63, Helsinki, FI-00014 University of Helsinki, Finland. ¹³⁸Department of Clinical Chemistry, Fimlab Laboratories and School of Medicine University of Tampere, Tampere, FI-33520, Finland. ¹³⁹Department of Medical Sciences, University Hospital, Uppsala, 75185, Sweden. ¹⁴⁰Transplantation laboratory, Haartman Institute, University of Helsinki, P.O. Box 21, Helsinki, FI-00014, Finland. ¹⁴¹National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina 27709, USA. ¹⁴²Ophthalmology, Massachusetts Eye and Ear, 243 Charles Street, Boston, Massachusetts 02114, USA. ¹⁴³Department of Statistics, University of Auckland, 303.325 Science Centre, Private Bag 92019, Auckland, 1142, New Zealand. ¹⁴⁴Department of Health, Functional Capacity and Welfare, National Institute for Health and Welfare, P.O. Box 30, Helsinki, FI-00271, Finland. ¹⁴⁵Department of Internal Medicine, University Hospital, Rue du Bugnon 44, Lausanne, 1011, Switzerland. ¹⁴⁶Human Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1HH, UK. ¹⁴⁷Division of Allergy and Clinical Immunology, Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21224, USA. ¹⁴⁸Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo, 108-8639, Japan. ¹⁴⁹Division of General Internal Medicine, Massachusetts General Hospital, 50 Staniford St, Boston, Massachusetts 02114, USA. ¹⁵⁰Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstr. 1, Neuherberg 85764, Germany. ¹⁵¹Institute of Human Genetics, Klinikum rechts der Isar, Technische Universität München, Ismaninger Str. 22, München 81675, Germany. ¹⁵²Molecular Epidemiology, QIMR Berghofer Medical Research Institute, 300 Herston Rd, Herston, Brisbane, Queensland 4006, Australia. ¹⁵³Genome Science Institute, Boston University School of Medicine, 72 East Concord Street, E-304, Boston, Massachusetts 02118, USA. ¹⁵⁴HUCH Heart and Lung center, Helsinki University Central Hospital, P.O. Box 340, Helsinki, FI-00029, Finland. ¹⁵⁵Pulmonary Center and Department of Medicine, Boston University School of Medicine, 72 E Concord St, Boston, Massachusetts 02118, USA. ¹⁵⁶Department of Medicine, University of Ibadan, Ibadan, Nigeria. ¹⁵⁷ICAMS, University of Glasgow, 126 University Way, Glasgow G12 8TA, UK. ¹⁵⁸Division of Epidemiology and Community Health, University of Minnesota, 1300 S 2nd Street, Minneapolis, Minnesota 55454, USA. ¹⁵⁹Centre for Global Health and Department of Public Health, School of Medicine, University of Split, Soltanska 2, 21000 Split, Croatia. ¹⁶⁰Obesity Research Unit, Research Programs Unit, Diabetes and Obesity, University of Helsinki, P.O.Box 63, Haartmaninkatu 8, FI-00014, Helsinki, Finland. ¹⁶¹International Centre for Circulatory Health, Imperial College London, London W2 1LA, UK. ¹⁶²Department of Genomics of Common Disease, School of Public Health, Imperial College London, London SW7 2AZ, UK. ¹⁶³Department of Cardiology and Cardiothoracic Surgery Hero DMC Heart Institute, Civil Lines, 141001, Ludhiana, India. ¹⁶⁴Department Public Health Sciences, University of Virginia School of Medicine, 3232 West Complex, Charlottesville, Virginia 22908, USA. ¹⁶⁵Department of Psychological & Brain Sciences, Indiana University Bloomington, 1101 E. 10th Street, Bloomington, Indiana 47405, USA. ¹⁶⁶Institute of Molecular Biology and Biochemistry, Medical University Graz,

- Harrachgasse 21, Graz, A-8010, Austria. ¹⁶⁷Science for Life Laboratory, Karolinska Institutet, Stockholm, SE-17121, Sweden. ¹⁶⁸University of Dundee, Kirsty Semple Way, Dundee DD2 4DB, UK. ¹⁶⁹Cardiovascular Health Research Unit, Division of Cardiology, University of Washington, 1730 Minor Ave, Suite 1360, Seattle, Washington 98101, USA. ¹⁷⁰Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, St. Stephen's Green, Dublin 2, Ireland. ¹⁷¹UMR INSERM U1122: IGE-PCV "Interactions Gène-Environnement en Physiopathologie Cardio-Vasculaire", INSERM, University of Lorraine, 30 Rue Lionnois, 54000 Nancy, France. ¹⁷²Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich 81377, Germany. ¹⁷³Department of Endocrinology, All India Institute of Medical Sciences, Ansari Nagar East, New Delhi, 110029, India. ¹⁷⁴National Heart and Lung Institute, Imperial College London, Du Cane Road, London W12 0NN, UK. ¹⁷⁵Department of Public Health Sciences, Stritch School of Medicine, Loyola University Chicago, Maywood, Illinois 60153, USA. ¹⁷⁶NIHR Leicester Cardiovascular Biomedical Research Unit, University of Leicester, Glenfield Hospital, Groby Road, Leicester LE3 9QP, UK. ¹⁷⁷Institute of Molecular and Cell Biology, University of Tartu, Riia 23, Tartu, 51010, Estonia. ¹⁷⁸Centre for Vascular Prevention, Danube-University Krems, 3500 Krems, Austria. ¹⁷⁹Diabetes Research Group, King Abdulaziz University, 21589 Jeddah, Saudi Arabia. ¹⁸⁰Department of Internal Medicine, Erasmus Medical Center, PO Box 2040, Rotterdam, 3000 CA, The Netherlands. ¹⁸¹Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland 21205, USA. ¹⁸²Finnish Lung Health Association, Sibeliuksenkatu 11 A 1, Helsinki, FI-00250, Finland. ¹⁸³Genetic Epidemiology, QIMR Berghofer Medical Research Institute, 300 Herston Rd, Herston, Brisbane, Queensland 4006, Australia. ¹⁸⁴Diabetes Unit, KEM Hospital and Research Centre, Rasta Peth, Pune, 411011, India. ¹⁸⁵Renal Unit, Department of Medicine, University of Verona, Piazzale A. Stefani 1, 37124 Verona, Italy. ¹⁸⁶Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio 44106, USA. ¹⁸⁷Department of Medicine, Stanford University, 300 Pasteur Drive, Stanford, California 94305, USA. ¹⁸⁸Department of Pediatrics, Perelman School of Medicine, The University of Pennsylvania, 3615 Civic Center Boulevard, Philadelphia, Pennsylvania 19104, USA. ¹⁸⁹Translational Gerontology Branch, National Institute on Aging, Baltimore, Maryland 21225, USA. ¹⁹⁰Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, No. 9 Edinburgh BioQuarter, 9 Little France Road, Edinburgh EH16 4UX, UK. ¹⁹¹Centre for Pharmacogenetics and Pharmacogenomics, Medical Research Institute, University of Dundee, Ninewells Hospital and School of Medicine, Dundee DD1 9SY, UK. ¹⁹²Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, 21589, Saudi Arabia. ¹⁹³Faculty of Medicine, Imperial College London, Charing Cross Campus, St Dunstan's Road, London W6 8RP, UK. ¹⁹⁴Department of Medicine, University of Leipzig, Leipzig 04103, Germany. ¹⁹⁵Institute of Preventive Medicine, Bispebjerg and Frederiksberg Hospital, The Capital Region, Copenhagen, 2000, Denmark. ¹⁹⁶Department of Epidemiology, University of Alabama at Birmingham, 1665 University Boulevard, Birmingham, Alabama 35294, USA. ¹⁹⁷Department of Psychiatry, University Medical Center Groningen, University of Groningen, P.O. Box 30.001, Groningen, 9700 RB, The Netherlands. ¹⁹⁸Epidemiology of diabetes, obesity and chronic kidney disease over the lifecourse, Inserm, CESP Center for Research in Epidemiology and Population Health U1018, 16 Avenue Paul Vaillant Couturier, 94807 Villejuif, France. ¹⁹⁹Dipartimento di Scienze Mediche, Catholic University of the Sacred Heart, Via G. Moscatti 31/34, 00168 Roma, Italy. ²⁰⁰Department of Biostatistics, University of Liverpool, Duncan Building, Daulby Stree, Liverpool L69 3GA, UK. ²⁰¹Department of General Practice and Primary Health Care, University of Helsinki, P.O. Box 20, University of Helsinki, Helsinki, FI-00014, Finland. ²⁰²Vasa Central Hospital, Sandviksgatan 2-4, Vasa, FI-65130, Finland. ²⁰³Folkhälsan Research Centre, PB 63, University of Helsinki, Helsinki, FI-00014, Finland. ²⁰⁴Unit of General Practice, Helsinki University Central Hospital, Haartmaninkatu 4, Helsinki, FI-00290, Finland. ²⁰⁵Neuro-Imaging Genetics, QIMR Berghofer Medical Research Institute, 300 Herston Rd, Herston, Brisbane, Queensland 4006, Australia. ²⁰⁶Cardiovascular Genetics Division, University of Utah, 420 Chipeta Way, Room 1160, Salt Lake City, Utah 84117, USA. ²⁰⁷Alzheimer Scotland Research Centre, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK. ²⁰⁸Department of Psychiatry, Erasmus Medical Center, PO Box 2040, Rotterdam, 3000 CA, The Netherlands. ²⁰⁹National Institute for Health and Welfare (THL), P.O.Box 30, Mannerheimintie 166, Helsinki, FI-00271, Finland. ²¹⁰Department of Kinesiology, Laval University, 2300 rue de la Terrasse, Quebec G1V 0A6, Canada. ²¹¹Department of Physiology and Biophysics, University of Mississippi Medical Center, 2500 N. State Street, Jackson, Mississippi 39216, USA. ²¹²Department of Clinical Physiology and Nuclear Medicine, University of Turku and Turku University Hospital, Turku, FI-20521, Finland. ²¹³Research Center of Applied and Preventive Cardiovascular medicine, University of Turku, Turku, FI-20521, Finland. ²¹⁴University of Groningen, University Medical Center Groningen, Department of Genetics, Hanzeplein 1, Groningen, 9700 RB, The Netherlands. ²¹⁵Department of Economics, Stockholm School of Economics, Box 6501, Stockholm, SE-113 83, Sweden. ²¹⁶Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Box 281, Stockholm, SE-171 77, Sweden. ²¹⁷Department of Genetics and Biostatistics, Washington University School of Medicine, 4444 Forest Park Boulevard, Saint Louis, Missouri 63108, USA. ²¹⁸Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland 21205, USA. ²¹⁹Department of Pediatrics, University of Oklahoma Health Sciences Center, 940 Stanton Young Boulevard, Oklahoma City, Oklahoma 73104, USA. ²²⁰Department of Pharmaceutical Sciences, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73104, USA. ²²¹Sidra Medical and Research Centre, Doha, Qatar. ²²²The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, New York 10029, USA. ²²³Genome Institute of Singapore, 60 Biopolis Street, #02-01 Genome, Singapore, 138672, Singapore. ²²⁴Program for Personalised and Genomic Medicine, Department of Medicine, University of Maryland School of Medicine, 685 Baltimore St. MSTF, Baltimore, Maryland 21201, USA. ²²⁵Geriatric Research and Education Clinical Center, Veterans Administration Medical Center, 685 W Baltimore MSTF, Baltimore, Maryland 21201, USA. ²²⁶BHF Glasgow Cardiovascular Research Centre, University of Glasgow, 126 University Place, Glasgow G12 8TA, UK. ²²⁷Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, 137 E. Franklin Street, Suite 306, Chapel Hill, North Carolina 27599, USA. ²²⁸Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology and Health Services, University of Washington, 1730 Minor Ave, Suite 1360, Seattle, Washington 98101, USA. ²²⁹Group Health Research Institute, Group Health Cooperative, 1730 Minor Ave, Suite 1360, Seattle, Washington 98101, USA. ²³⁰Imperial College Healthcare NHS Trust, Imperial College London, Praed Street, London W2 1NY, UK. ²³¹Population Health Research Institute, St George's, University of London, Cranmer Terrace, London SW17 0RE, UK. ²³²Steno Diabetes Centre, Niels Steensens Vej 2, Gentofte, 2820, Denmark. ²³³Translational Laboratory in Genetic Medicine (TLGM), Agency for Science, Technology and Research (A*STAR), 8A Biomedical Grove, 138648, Singapore. ²³⁴Centre for Population Health Sciences, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, UK. ²³⁵Institutes for Neurogenetics and Integrative & Experimental Genomics, University of Lübeck, Lübeck 23562, Germany. ²³⁶School of Biotechnology, Jawaharlal Nehru University, New Delhi 110067, India.

*These authors contributed equally to this work.



Extended Data Figure 1 | Forest plot for cognitive ability (g). Individual sub-cohort estimates of effect size and the 95% confidence interval are plotted. Sub-cohorts are ordered from top to bottom according to their weight in the meta-analysis, so larger or more homozygous cohorts appear towards the top. The scale of β_{FROH} is in intra-sex standard deviations. The meta-analytical estimate is displayed at the bottom. Sub-cohort names follow the conventions detailed in Supplementary Table 6 and the Supplementary Table 11 legend. Sample sizes, effect sizes and P values for association are given in Table 1. This trait was rank-transformed.

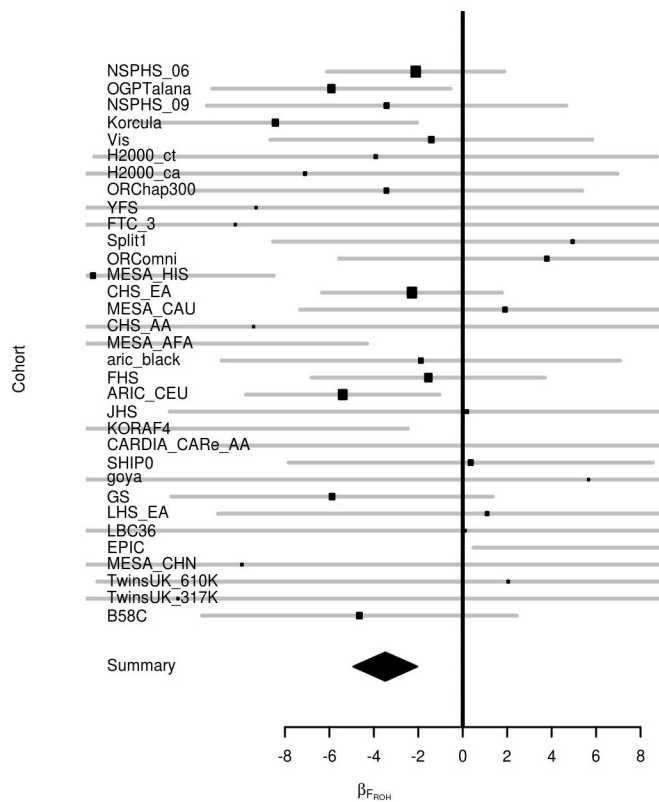


Extended Data Figure 2 | Forest plot for educational attainment. Individual sub-cohort estimates of effect size and the 95% confidence interval are plotted. Sub-cohorts are ordered from top to bottom according to their weight in the meta-analysis, so larger or more homozygous cohorts appear towards the top. The scale of β_{FROH} is in intra-sex standard deviations. The meta-analytical estimate is displayed at the bottom. Sub-cohort names follow the conventions detailed in Supplementary Table 6 and the Supplementary Table 11 legend. Sample sizes, effect sizes and *P* values for association are given in Table 1.

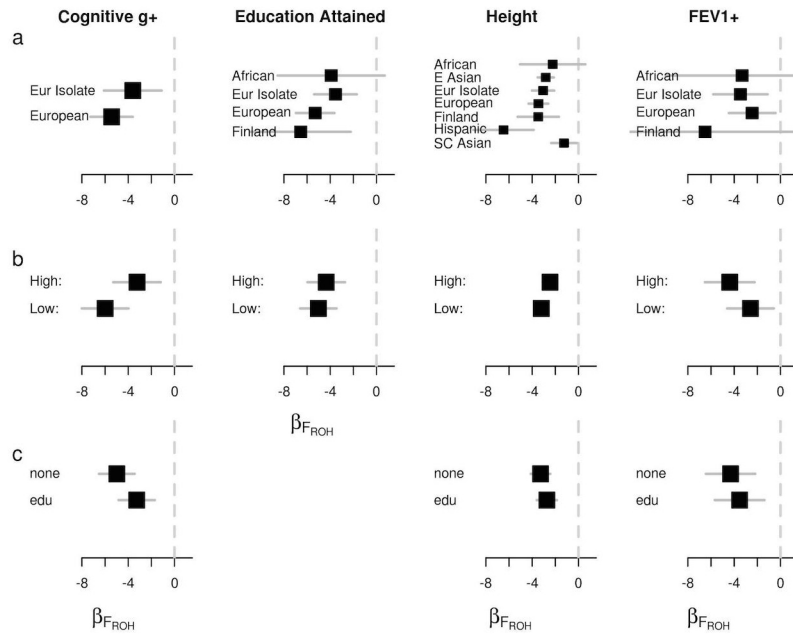


Extended Data Figure 3 | Forest plot for height. Individual sub-cohort estimates of effect size and the 95% confidence interval are plotted. Sub-cohorts are ordered from top to bottom according to their weight in the meta-analysis, so larger or more homozygous cohorts appear towards the top. The scale of

β_{FROH} is in intra-sex standard deviations. The meta-analytical estimate is displayed at the bottom. Sub-cohort names follow the conventions detailed in Supplementary Table 6 and the Supplementary Table 11 legend. Sample sizes, effect sizes and P values for association are given in Table 1.

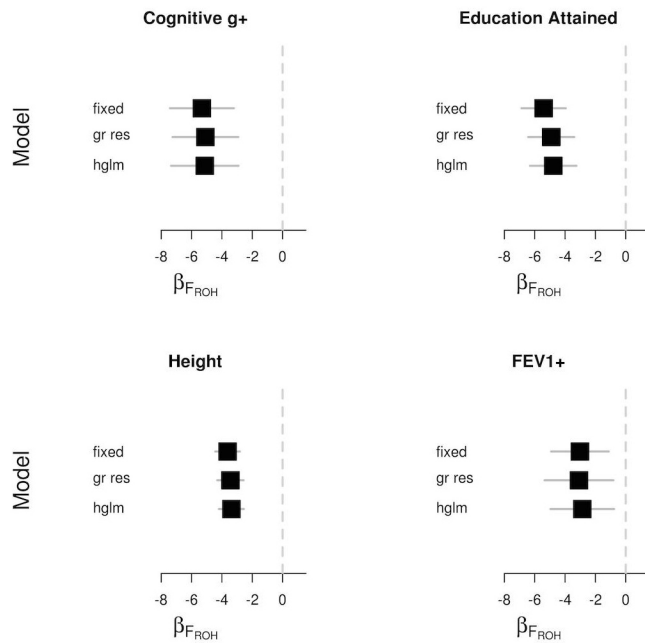


Extended Data Figure 4 | Forest plot for forced expiratory lung volume in one second. Individual sub-cohort estimates of effect size and the 95% confidence interval are plotted. Sub-cohorts are ordered from top to bottom according to their weight in the meta-analysis, so larger or more homozygous cohorts appear towards the top. The scale of β_{FROH} is in intra-sex standard deviations. The meta-analytical estimate is displayed at the bottom. Sub-cohort names follow the conventions detailed in Supplementary Table 6 and the Supplementary Table 11 legend. Sample sizes, effect sizes and P values for association are given in Table 1. This trait was rank-transformed.

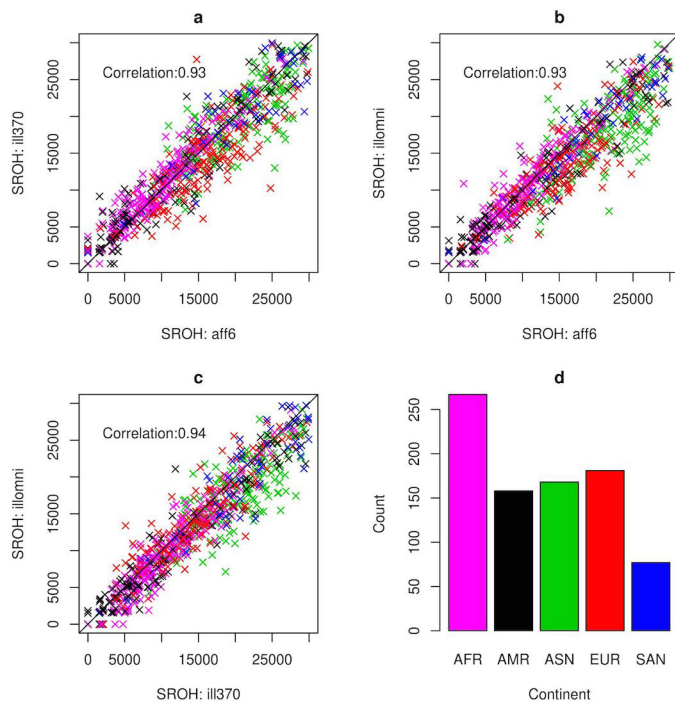


Extended Data Figure 5 | Signals of directional dominance are robust to stratification by geography or demographic history or inclusion of educational attainment as covariate. **a**, Cohorts are divided by continental biogeographic ancestry (African (15 sub-cohorts), East Asian (5), South and Central Asian (SC Asian; 10), Hispanic (3)), with Europeans being divided into Finns (13), other European isolates (self-declared, 23), and (non-isolated) Europeans (90). Meta-analysis was carried out for all subsets with 2,000 or more samples available. Sample numbers are as follows: cognitive g, Eur isolate, 6,638; European, 44,153; educational attainment, African 4,811; Eur isolate, 8,032; European, 55,549; Finland 9,068; height, African, 21,500; E Asian, 30,011; Eur isolate, 23,116; European, 228,813; Finland, 30,427; Hispanic, 5,469; SC Asian, 13,523; FEV1, African, 6,604; Eur isolate, 4,837; European, 49,223; Finland, 2,340. β_{FROH} is consistent across geography and in both isolates and more cosmopolitan populations. **b**, Cohorts were divided into high and low ROH strata of equal power and meta-analysis repeated – the effects are consistent

across strata for all four traits. The mean SROH for the high and low strata, respectively, are 13.4 and 4.3 Mb for cognitive g; 28.1 and 5.1 Mb for educational attainment; 31.9 and 10.8 Mb for height; and 41.4 and 4.5 Mb for FEV1. **c**, To assess the potential for socio-economic confounding, where available, educational attainment was included in the regression model (edu) and compared to a model without educational attainment (none) in the same subset of cohorts. The signals reduce slightly when the education covariate is included; the analysis is not possible for educational attainment as a trait. For cognitive g, numbers of subjects are 36,847 and 36,023; for height 131,614 and 120,945; and for FEV1, 15,717 and 15,425, for edu and none, respectively. The numbers differ because of missing individual educational data within cohorts. Plus signs indicate that the phenotype was rank-transformed. Trait units are intra-sex standard deviations and the genomic measure is unpruned SROH. Subset estimates of effect size for FROH and the 95% confidence are plotted.



Extended Data Figure 6 | Signals of directional dominance are robust to model choice. Meta-analytical estimates of effect size and standard errors are plotted for various models. Fixed, no mixed modelling was used; gr res, GRAMMAR+ residuals were fitted; hglm, full hierarchical generalized linear mixed model was used. Plus signs indicate that the phenotype was rank-transformed. 15,355 subjects were used for cognitive g, 36,060 for educational attainment, 89,112 for height and 15,262 for FEV1.



Extended Data Figure 7 | Correlation in SROH for different genotyping arrays using HapMap populations. **a–c**, x and y axes show SROH from 0–30 Mb. ill370, Illumina CNV370; aff6, Affymetrix6; ilomni, Illumina OmniExpress. The graphs are shown for the specific PLINK call parameters used. **d**, Sample numbers per continent are presented in a bar chart. AFR, African; AMR, mixed American; ASN, East Asian; EUR, European; SAN, South Asian. Only samples with SROH below 30 Mb are plotted, to be conservative to the effect of outliers, which have very strongly correlated estimates of SROH ($r = 0.96$ – 0.97 for comparisons including such very homozygous individuals). In these plots, the correlation between SROH called by the two arrays, $r = 0.93$ – 0.94 .

Extended Data Table 1 | Continental ancestry of cohorts participating in each trait study.

	African	East Asian	European	Hispanic	S/C Asian	All
BMI	21689/15	29009/5	279400/117	7836/3	13464/10	351398/150
Cognitive <i>g</i>	1539/1	NA/NA	49559/22	-	-	51098/23
Diastolic BP	17074/12	24200/5	204742/85	7284/3	12876/9	266176/114
Education Attained	4811/4	NA/NA	79576/42	-	338/1	84725/47
Fasting Insulin	6895/8	1603/1	72006/49	-	6303/5	86807/63
FEV1	6604/5	617/1	58089/27	825/1	-	66135/34
FEV1/FVC	6565/5	616/1	57888/27	822/1	-	65891/34
FP Glucose	8942/9	1615/1	122368/74	1938/1	6921/5	141784/90
HbA1c	6629/4	694/1	92732/31	4038/2	7509/4	111602/42
HDL Cholesterol	15099/13	10478/5	215621/92	4426/3	12508/9	258132/122
Height	20300/14	30011/5	281369/114	5469/2	13523/10	350672/145
LDL Cholesterol	13375/11	2503/2	172245/77	4340/3	11186/8	203649/101
Systolic BP	17023/12	24424/5	205253/85	7225/3	12859/9	266784/114
Total Cholesterol	15130/13	20187/5	209421/91	4491/3	11674/8	260903/120
Triglycerides	13886/12	2542/2	181526/84	2745/2	10688/7	211387/107
Waist-hip ratio	8182/7	2549/2	171753/73	1446/1	12598/9	196528/92

The first number in each cell is the number of participants with that continental ancestry. The second number is the number of sub-cohorts. S/C Asian, South and Central Asian.

Parent–progeny sequencing indicates higher mutation rates in heterozygotes

Sihai Yang^{1*}, Long Wang^{1*}, Ju Huang^{1*}, Xiaohui Zhang¹, Yang Yuan¹, Jian-Qun Chen¹, Laurence D. Hurst² & Dacheng Tian¹

Mutation rates vary within genomes, but the causes of this remain unclear¹. As many prior inferences rely on methods that assume an absence of selection, potentially leading to artefactual results², we call mutation events directly using a parent–offspring sequencing strategy focusing on *Arabidopsis* and using rice and honey bee for replication. Here we show that mutation rates are higher in heterozygotes and in proximity to crossover events. A correlation between recombination rate and intraspecific diversity is in part owing to a higher mutation rate in domains of high recombination/diversity. Implicating diversity *per se* as a cause, we find an ~3.5-fold higher mutation rate in heterozygotes than in homozygotes, with mutations occurring in closer proximity to heterozygous sites than expected by chance. In a genome that is a patchwork of heterozygous and homozygous domains, mutations occur disproportionately more often in the heterozygous domains. If segregating mutations predispose to a higher local mutation rate, clusters of genes dominantly under purifying selection (more commonly homozygous) and under balancing selection (more commonly heterozygous), might have low and high mutation rates, respectively. Our results are consistent with this, there being a ten times higher mutation rate in pathogen resistance genes, expected to be under positive or balancing selection. Consequently, we do not necessarily need to evoke extremely weak^{1,2} selection on the mutation rate to explain why mutational hot and cold spots might correspond to regions under positive/balancing and purifying selection, respectively^{3,4}.

To determine mutation rates, we selected two purebred parents in both *Arabidopsis* (strains Col and Ler) and rice (strains 9311 and PA64s) (Fig. 1). We selfed each and sequenced both parents (P_0) and progeny (P_1). In addition, we crossed to generate intraspecific F_1 heterozygotes. A single heterozygous F_1 seed in each species was selfed to generate multiple F_2 progeny. By comparing sequences between F_2 seeds we could determine the $F_1 \rightarrow F_2$ mutation rate. While direct sequencing of genomes is the best way to detect *de novo* mutations^{5,6}, the error rate is high. We negate this by having several lines of quality control (Extended Data Fig. 1a). First, we sequenced multiple independent DNA extractions from the same individual or inbred progeny of the individual, permitting a mutation to be called only if replicates agree. In practice, a mutation called in one extract from a given plant was always found in replicates. In addition, we use a consensus approach, comparing each focal individual against all other relevant samples⁷. For example, a presumptive mutation in an F_2 progeny must be both called within a ‘mutated’ sample and not called in both the sequenced parental genomes and all other F_2 progeny. These criteria ensure that the mutation must have arisen sometime after bolting in the F_1 , as all other F_2 progeny share the same F_1 parent. To call a mutation we additionally require high sequence quality (score ≥ 30 ; detail in Supplementary Table 1) and high coverage ($>6,000\times$ for the sample cohorts and $>40\times$ for each sample) with at least five or more reads that must include both the forward and

reverse reads. This approach is robust against sequencing or alignment errors in the reference genome⁷. False positive rates are negligible, while false negative rates are low (Methods).

In *Arabidopsis*, 237 base mutations and 67 small insertion/deletions (indels) were detected in the 26 progeny of selfing *Arabidopsis* parents (P_0 to P_1) and 67 F_2 plants (F_1 to F_2) from the Col \times Ler cross (Fig. 1, Table 1 and Supplementary Tables 2, 3). To assess their reliability, several strategies were applied. First, Sanger sequencing confirmed 100% of 112 sampled base mutations and 100% of 43 sampled indels present in F_2 plants. Confirmation requires that the mutation be present in the focal individual and absent in both parental genomes. Second, the 32 sequenced F_4 plants, derived from two F_2 plants (c52 and c64 with 10 mutations observed) (Fig. 1 and Supplementary Table 3), confirmed 100% of these F_2 mutations at a frequency of ~73% (slightly higher than the expected 62.5%). Third, we randomly sampled 4–8 F_3 plants from each of the 21 sampled F_2 plants and subjected these F_3 plants to Sanger sequencing. This confirmed 99 out of 100 sampled base mutations and 24 out of 26 indels present in F_3 plants.

Comparison with previous estimates suggests that our method is robust. We estimate a rate of 7.4×10^{-9} base mutations per generation per site in homozygous individuals (that is, $P_0 \rightarrow P_1$), similar to the previous estimate of 7.0×10^{-9} from mutation-accumulation Col lines⁸. As typically reported, we observe more transitions than transversions (Supplementary Table 4 and Extended Data Fig. 2a) and that mutations are disproportionately common at GC-rich nucleotide triplets (Supplementary Table 5). The ratio of point mutations to indels (3.9) is in line with previous estimates (3.11–5.8) (refs 8, 9). Mutations in Col \times Ler F_1 hybrids are as likely to occur on the Ler genome as on the Col genome ($\chi^2 = 1.4$, degrees of freedom (d.f.) = 1, $P = 0.23$).

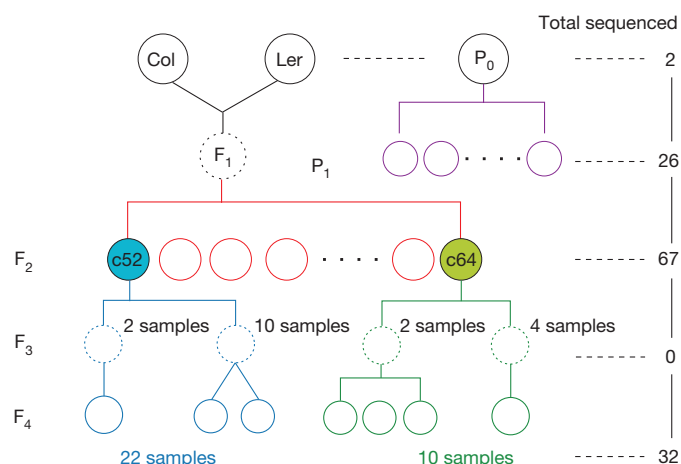


Figure 1 | Pedigree relationship of the sequenced *Arabidopsis* samples. The number of circles with solid lines denotes how many samples from each generation are sequenced; for example, the sequenced samples from c52 = $(2 \times 1) + (10 \times 2) = 22$.

¹State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing 210023, China. ²The Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK.

*These authors contributed equally to this work.

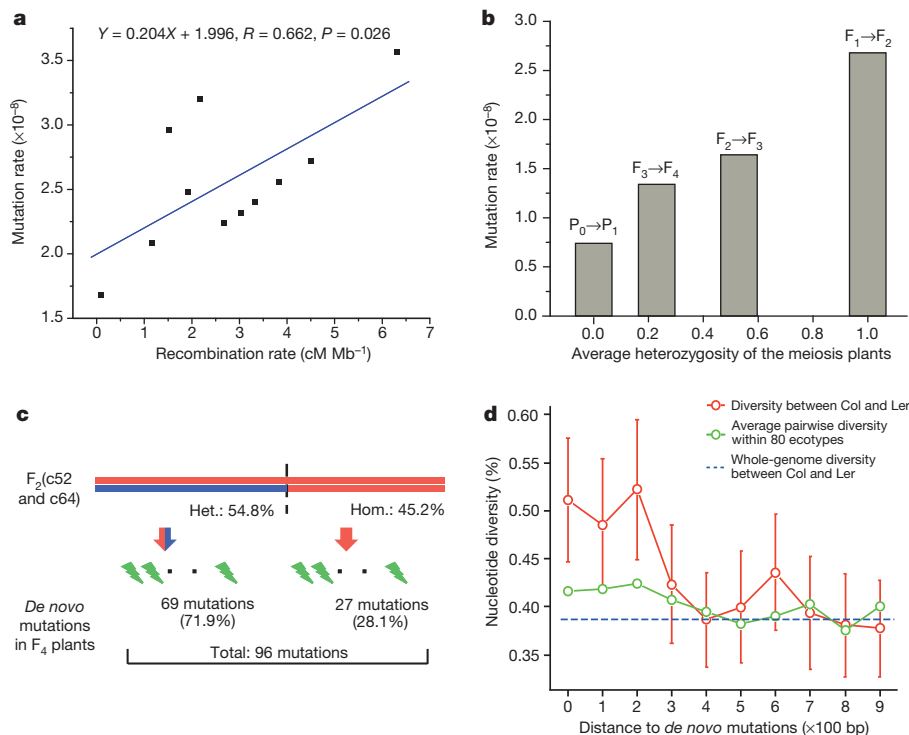
Table 1 | Number of spontaneous mutations per meiosis in the *Arabidopsis* genome

Genotypes of the plants with meioses	Sequenced samples	Base mutations (average mutations per sample)			Indel mutations (average mutations per sample)		
		Non-repeat regions	Repeat regions	Total (average)	Non-repeat regions	Repeat regions	Total (average)
Homozygotes ($P_0 \rightarrow P_1$)	26 P_1 plants	18 (0.69)	5 (0.19)	23 (0.88)	6 (0.29)	2 (0.08)	8 (0.31)
Heterozygotes ($F_1 \rightarrow F_2$)	67 F_2 plants	164 (2.45)	50 (0.75)	214 (3.19)	49 (0.673)	10 (0.15)	59 (0.88)
Heterozygotes ($F_2 \rightarrow F_4$)	32 F_4 plants	Specific	52 (1.62)	21 (0.66)	73 (2.28)	11 (0.34)	2 (0.06)
		Shared	4	5	9	1 (0.03)	0
	Average mutations/sample of $F_2 \rightarrow F_3$			(1.92)	ND		
	Average mutations/sample of $F_3 \rightarrow F_4$			(1.61)	ND		

The indel sizes range from 1 to 27 bp (2.91 on average; see Supplementary Table 8). The calculation of average mutations in the meiosis from $F_2 \rightarrow F_3$ and $F_3 \rightarrow F_4$ is described in Extended Data Fig. 1b. ND, not determined because the number of indels is too small to calculate the average indel mutations per sample of $F_2 \rightarrow F_3$ or $F_3 \rightarrow F_4$.

We note one deviation from null expectation, this being a higher density of mutations in *Arabidopsis* non-coding compared with coding regions, which cannot be accounted for in terms of differences in trinucleotide content (Supplementary Table 6). This suggests either underestimation of the mutation rate in coding sequences, possibly due to purifying selection, or a lower mutation rate in the transcribed sequence, possibly owing to transcription-coupled repair. A selectionist explanation predicts an increased relative frequency of indels that are multiples of three long in the coding sequence. Even using a one-tailed test, we find no evidence for this. Of 81 indels, 62 and 12 are not multiples of three and are outside and inside the coding sequence, respectively, while 5 (outside) and 2 (inside) are multiples of three (Fisher's exact test, one-tailed $P = 0.35$).

Population-wide intragenomic diversity is commonly reported to be higher in genomic domains of high crossing-over¹⁰, which we also see in *Arabidopsis* (Extended Data Fig. 3a). This is typically ascribed to reduced selective interference between physically close alleles in domains of high recombination¹⁰. However, it might reflect a tendency for regions with high recombination rates to also be domains with high mutation rates, possibly because recombination is mutagenic^{11–16}. Dissecting the chromosomes into 1 Mb non-overlapping regions, we indeed find a positive correlation between mutation rates and the rates of crossover events in the 67 *Arabidopsis* F_2 and 32 F_4 plants (Fig. 2a). This is consistent with the possibility either that recombination is mutagenic or that both mutation and recombination preferentially occur in high diversity domains.

**Figure 2 | Patterns of diversity, recombination and *de novo* mutation.**

a, Relationship between the mutation and recombination rate. When the chromosomes are dissected into 1 Mb non-overlapping regions, the recombination rate (cM Mb^{-1}) and mutation number per Mb can be obtained for each of them. When ranked then sorted by the recombination rates, the mean mutation rate per recombinational class was obtained. Line is standard regression (for relationship between recombination and diversity see Extended Data Fig. 3a). **b**, **c**, Variation in the mutation rates as a function of heterozygosity proportion during meiosis. Detailed calculation of mutation rates of $F_3 \rightarrow F_4$, $F_2 \rightarrow F_3$ is shown in Extended Data Fig. 1b. The number of mutations was counted separately in the regions of F_4 plants derived from heterozygous or homozygous regions of F_2 plants, respectively. **c**, The

percentages in brackets reflect the proportion of the genome that is heterozygous (Het.) or homozygous (Hom.). **d**, Relationship between nucleotide diversity and the distance to the *de novo* mutations. Window 0 in x -axis is the 2×100 bp sequence surrounding the position of any given *de novo* mutation and 1–9 is 100–900 bp away from the mutation on both sides. For each window of 2×100 bp sequence, the average diversity is calculated. The red circles denote the diversity between Col and Ler—that is, heterozygosity of parents—the green circles are the average diversity among 80 *Arabidopsis* populations at the same windows³⁰, and the blue dashes are the average genomic diversity (0.39%) between the two parental genomes (Col and Ler). Error bars, mean \pm standard error of the mean. Test for difference in slope, $Z = 3.08$, $P = 0.002$.

Given the very high intragenomic variation in crossover rates seen in honey bees, we examined the possibility that mutation happens more commonly in the vicinity of crossovers by examining *de novo* mutations in 46 honey bee genomes. In this species too, intraspecific diversity is correlated with the crossing-over rate¹⁷. Mutagenic effects of crossing over are thought to occur within 2 kb of the break point¹⁶. Of 35 mutations, 2 mutations occurred within a 2 kb distance of a crossover breakpoint¹⁷ ($P = 0.0012$ with 10,000 randomizations; Extended Data Fig. 2b). Thus, in this genome also, new mutations occur in proximity to crossover events more often than expected by chance. We estimate the per-genome mutation rate of a diploid queen to be 9.0×10^{-9} (6.8×10^{-9} for base substitution and 2.2×10^{-9} for indels).

Although mutagenic repair may be acting in the immediate vicinity of a double-strand break (DSB)^{11–16}, a higher rate of both recombination and mutation (mechanistically uncoupled) in domains of high diversity provides an alternative explanation for the correlation between mutation and recombination. That intraspecific diversity in *Arabidopsis* correlates with between-species divergence (Extended Data Fig. 3b) is consistent with either possibility. A possible coupling between mutation and intragenomic diversity (that is, heterozygosity) could be found if heterozygosity causes an increase in the mutation rate¹⁸. We test this by comparing progeny derived from heterozygous and homozygous parents in our two plant species. The point mutation rate (2.68×10^{-8}), as assayed from analysis of the F_2 progeny of heterozygous F_1 *Arabidopsis*, is ~ 3.6 -fold higher than that in the homozygous progeny of the selfed parents (two-tailed Brunner–Munzel (BM) test, $P = 3.64 \times 10^{-8}$). Similarly, the indel mutation rate in intergenic regions in heterozygote F_2 progeny is ~ 2.8 -fold higher than that in homozygotes (Table 1; two-tailed BM test, $P = 0.0012$). The same pattern is seen in rice lines with 3.4-fold higher mutation rates in heterozygotes (3.2×10^{-9} and 1.1×10^{-8} per site per meiosis in homozygotes and heterozygotes, respectively) (Table 2; two-tailed BM test, $P = 0.0028$). Analysis of 158 *Arabidopsis* point mutations in which Col, Ler and *A. lyrata* have the same state before mutation (and are thus unlikely to be hypermutagenic), suggests that Col–Ler F_1 has a 5.02-fold larger mutation rate than the selfed Col or Ler parents ($P_0 \rightarrow P_1$) (BM test, $P = 1.02 \times 10^{-7}$).

The possibility that the degree of heterozygosity predicts the mutation rate can be further tested. Compared with F_2 , a reduced mutation rate is expected in F_3 or F_4 selfed plants because the heterozygous regions will reduce by one-half each generation. We identified 86 mutations detected in only one of the 32 F_4 plants, comprising 73 base and 13 indel mutations, giving a base mutation rate of 1.34×10^{-8} in F_4 plants, inherited from 18 F_3 of 2 F_2 plants (c52 and c64 in Fig. 1), and 1.60×10^{-8} in F_3 plants (methods were as described previously⁸; see Fig. 2b and Extended Data Fig. 1b for details). This ordering is as expected under the assumption that heterozygosity predicts mutation rates.

Were mutations easier to call in heterozygous regions, the observations described earlier may be artefactual. To address this, we

considered mutations from the F_1 to the F_2 generation. In some genomic domains the F_2 preserves the heterozygosity of the F_1 (which is uniformly heterozygous during meiosis), while in some genomic locations the F_2 is homozygous. If artefact were to explain higher call rates in the heterozygous regions, we would expect more mutants to be called in the F_2 heterozygous domains. We do not observe this (153 mutations in the 54% heterozygous domains, 120 in homozygous domains, expected 146.5 and 126.5, respectively, allowing for trinucleotide content; χ^2 with Yates correction = 0.53, d.f. = 1, $P = 0.47$).

These results may reflect either (1) a tendency for heterozygotes to have genome-wide disruption of the mechanisms that prevent mutation (for example, owing to disruption of co-adapted heteromers), this being dependent on the proportion of the genome that is heterozygous; or (2) a genomically local effect of heterozygosity on the mutation rate. If the latter is the case, in genomes that are stratified into heterozygous and homozygous blocks the mutation rate should be higher in heterozygous domains. We observe this. There are 69 mutations in the regions of F_4 plants derived from heterozygous regions of both c64 (48% heterozygous blocks) and c52 (61% heterozygous blocks), compared with 27 in the regions of F_4 plants from homozygous regions of F_2 plants. Allowing both for the proportion of the genome covered and for differences in trinucleotide content, there is an excess in domains of heterozygosity (expected 52.3 and 43.7; χ^2 with Yates correction = 11.02, d.f. = 1, $P < 0.001$; Fig. 2c). Analysis of non-hypermutable sites confirms the same (χ^2 with Yates correction = 6.11, d.f. = 1, $P = 0.01$).

A more conservative version of this test examines the 96 mutations accumulated in the F_4 since the F_2 plants, in the 7% of the genome remaining heterozygous in the F_4 . While such regions have a longer history of heterozygosity, many of the domains homozygous in the F_4 were heterozygous in the F_3 . Nonetheless, we observe more mutations than expected in the heterozygous spans (heterozygous span: observed 13, expected 6.85; homozygous span: observed 83, expected 89.15; χ^2 with Yates correction = 5.02, d.f. = 1, $P = 0.02$). Analysis of non-hypermutable sites confirms this (χ^2 with Yates correction = 4.13, d.f. = 1, $P = 0.04$). These data support the notion that heterozygosity is associated with a local increase in the mutation rate.

If heterozygosity is causative, we might expect mutational events to be close to heterozygous sites in the parents, whereas sites that are polymorphic in the population but not in the parents need not be in especially close proximity to mutations. We find that parental heterozygous sites are significantly closer to mutational sites than expected (Fig. 2d, red circles). There are a total of 273 mutations raised from $F_1 \rightarrow F_2$. The median distance of the *de novo* mutation to a heterozygous site is 167 bp (0 to 32,694 bp), significantly smaller than the expected median distance with a random null (10,000 randomizations, expected median = 207 bp; $P = 0.05$). Of those mutations, 113 are within 100 bp of heterozygous sites, significantly more than expected by chance (10,000 randomizations, expected number = 93, $P = 0.005$). As also expected, the level of diversity within the parents surrounding mutation sites is higher than the genome average (0.39% between two parents). By contrast, population polymorphism shows no such trend (Fig. 2d and Extended Data Fig. 3c). The different patterns are consistent with local heterozygosity in the parent being causative, but a bias towards heterozygosity and mutation to both being intergenic might provide an alternative rationale.

On a broader scale, if we bin the genome into windows of 1 Mb, we find a correlation between mutation rate and intraspecific diversity (Spearman's $\rho = 0.76$, $P = 0.0059$), suggestive of wide-scale mutational domains that impact on levels of polymorphism. If heterozygosity causes mutations, such domains might be self-reinforcing, but the correlation alone is not evidence for this. Such an autocatalytic process suggests that both the highly polymorphic regions within a species and the species with higher rates of outcrossing could have higher mutation rates, compared with the conserved regions or self-crossing species, respectively. A number of studies indicate that the mating system (outcrossing or selfing), affects the mutation rate¹⁹ and that mutations occur

Table 2 | Numbers of spontaneous mutations (new SNPs) per meiosis in rice F_2 plants

Samples		SNPs	Indels	Samples	SNPs	Indels
Homozygotes	9311-1	0	1	PA64s-1	1	0
	9311-2	1	1	PA64s-2	1	2
	9311-3	3	0	Average	1.2	0.8
Heterozygotes (F_2 seeds)	F2_22	5	2	F2_32	3	1
	F2_23	3	0	F2_56	1	1
	F2_24	4	2	F2_88	6	2
	F2_25	11	1	F2_89	6	0
	F2_27	2	1	F2_90	1	0
	F2_30	3	0	Average	4.09	0.91

Two rice samples, PA64s-3 and F2_29, were removed owing to their low sequencing quality in one of the independent sequencings. The base substitution mutation rate is 3.2×10^{-9} and 1.1×10^{-8} per site per meiosis in the homozygous and heterozygous rice genomes, respectively. SNPs, single nucleotide polymorphisms.

near pre-existing diversity²⁰, particularly near indels²¹. However, as selfers and asexuals can retain linkage disequilibrium between mutator alleles and mutations, genome-wide selection on the mutation rate could confound between-species comparisons. More generally, understanding between-species variation is likely to be difficult owing to expected covariation with parameters that are difficult to determine, such as the effective population size.

It has been observed^{3,4} that genomic hot and cold spots of indirectly inferred mutations accord with domains of genes putatively under strong purifying selection (mutational cold spots) and positive or balancing selection (mutational hot spots). However, this observation might be an artefact of indirect methods to detect mutations: putatively neutral mutations in genes under strong purifying selection might be purged by selection if not neutral, causing an underestimation of mutation rates². Our sequencing strategy largely avoids this problem. Nonetheless, we find evidence that genes expected to be under positive/balancing selection have high mutation rates. In *Arabidopsis*, a total of 68 base mutations and 14 small indels occurred in coding sequences either as synonymous (21) or non-synonymous or frameshift mutations (59; Supplementary Table 6). Remarkably, 12 mutations are found in very few highly diversified gene families only and hence are prime targets of positive/balancing selection (Supplementary Table 7). Particular hot spots include nine *LRR*-encoding (associated with pathogen resistance) and three *F-box* genes, for which observed numbers greatly exceed the expected 0.89 (~10-fold higher) and 0.68 (~4.4-fold higher) mutations per family in these *Arabidopsis* F₂ plants, respectively (Supplementary Table 7). Of the 17 coding mutations previously reported⁸, one *NBS-LRR* gene (*AT1G59780*) and one *LRR-RLK* were detected (Supplementary Table 7), suggesting that this result is repeatable. *LRR*-encoding and *F-box* genes have a lower GC content than average (42.6% and 42.1%, respectively, versus mean of 44%), suggesting that this is not owing to underlying nucleotide mutability.

While at first sight a higher mutation rate in genes associated with pathogen resistance (and positive/balancing selection more generally) makes sense in terms of selection acting on the mutation rate^{3,18,22,23}, such modifiers of the mutation rate acting locally will have such weak selection on them that such an explanation makes little theoretical sense^{1,2}, especially when population sizes are small. Our results suggest a resolution of this paradox: genes subject to balancing selection will have a higher chance of being heterozygous, thus increasing the local mutation rate. That is to say, the selected variants could themselves be the modifiers of the mutation rate and hence their increase in frequency is attributed not to weak selection on the mutation rate, but strong selection on the direct phenotypic effects of some of the mutations.

We do not presume that heterozygosity is the only possible coupling between mutation and 'non-essentiality'. Indeed, an explanation based on heterozygosity is not of obvious relevance to bacteria. The effect we observed suggesting correlation between DSB events and mutation might, however, be more general. Indeed, in bacteria, DSB events can be mutagenic²⁴ and one need only hypothesize a coincidence between such recombination and non-essentiality, as seen in several eukaryotes²⁵, to provide an alternative explanation for hot and cold mutational spots. More immediate effects of transcription-coupled repair/mutation might also be of relevance.

While we make no attempt to investigate the underlying mechanism, we can speculate as to how heterozygosity might promote mutation. Several suggestions have been made¹⁸, to which we add a possible coupling with poor pairing during meiosis, as an immediate consequence of heterozygosity, especially for indels, may be poor pairing quality or failure of homology search. Poor pairing might be mutagenic because physically exposed regions are more likely to proceed to Spo11-mediated DSBs^{26,27}, repair of which is thought to be prone to error²⁸. Similarly, the DNA damage response protein MDC1 promotes accumulation of the sensor kinase ATR on unsynapsed chromosomes

and chromatin loops in mammals²⁹. Extended Data Fig. 2c illustrates such a possible mechanism. In this region there are great differences in both length (47 kb versus 48 kb) and diversity (~10% between AT3G23110 and AT3G23120) between Col and Ler (or homologous chromosomes in the F₁).

A caveat about our results is that the extent of size difference between Col and Ler is such that it may be unrepresentative of what normally happens in meiosis. Nonetheless, the poor-pairing model has the advantage that it might also explain the domains of higher mutation rate in homozygous Col⁸. During meiosis in homozygotes, repeating sequences (including clusters of homologous genes) can find homologous sequences at non-orthologous sites (ectopic recombination) and so force unpaired regions between homologous chromosomes. We analysed the repeat sequences in and around our 145 and the previously found^{8,9} 42 mutation-bearing genes in homozygotes. Consistent with expectations, 84.8% and 85.7% of these genes, including the gene *AT1G59780*, are located in repeat sequences or homologous gene clusters (Supplementary Table 7).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 April 2014; accepted 11 June 2015.

Published online 15 July 2015.

- Hodgkinson, A. & Eyre-Walker, A. Variation in the mutation rate across mammalian genomes. *Nature Rev. Genet.* **12**, 756–766 (2011).
- Chen, X. & Zhang, J. No gene-specific optimization of mutation rate in *Escherichia coli*. *Mol. Biol. Evol.* **30**, 1559–1562 (2013).
- Chuang, J. H. & Li, H. Functional bias and spatial organization of genes in mutational hot and cold regions in the human genome. *PLoS Biol.* **2**, e29 (2004).
- Martincoren, I., Seshasayee, A. S. N. & Luscombe, N. M. Evidence of non-random mutation rates suggests an evolutionary risk management strategy. *Nature* **485**, 95–98 (2012).
- Roach, J. C. *et al.* Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **328**, 636–639 (2010).
- Wang, J., Fan, H. C., Behr, B. & Quake, S. R. Genome-wide single-cell analysis of recombination activity and *de novo* mutation rates in human sperm. *Cell* **150**, 402–412 (2012).
- Sung, W., Ackerman, M. S., Miller, S. F., Doak, T. G. & Lynch, M. Drift-barrier hypothesis and mutation-rate evolution. *Proc. Natl Acad. Sci. USA* **109**, 18488–18492 (2012).
- Ossowski, S. *et al.* The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* **327**, 92–94 (2010).
- Jiang, C. *et al.* Environmentally responsive genome-wide accumulation of *de novo* *Arabidopsis thaliana* mutations and epimutations. *Genome Res.* **24**, 1821–1829 (2014).
- Cutter, A. D. & Payseur, B. A. Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature Rev. Genet.* **14**, 262–274 (2013).
- Lercher, M. J. & Hurst, L. D. Human SNP variability and mutation rate are higher in regions of high recombination. *Trends Genet.* **18**, 337–340 (2002).
- Magni, G. E. & Borstel, R. C. V. Different rates of spontaneous mutation during mitosis and meiosis in yeast. *Genetics* **47**, 1097–1108 (1962).
- Perry, J. & Ashworth, A. Evolutionary rate of a gene affected by chromosomal position. *Curr. Biol.* **9**, 987 (1999).
- Pratto, F. *et al.* Recombination initiation maps of individual human genomes. *Science* **346**, 1256442 (2014).
- Ratray, A., Santoyo, G., Shafer, B. & Strathern, J. N. Elevated mutation rate during meiosis in *Saccharomyces cerevisiae*. *PLoS Genet.* **11**, e1004910 (2015).
- Arbeithuber, B., Betancourt, A. J., Ebner, T. & Tiemann-Boege, I. Crossovers are associated with mutation and biased gene conversion at recombination hotspots. *Proc. Natl Acad. Sci. USA* **112**, 2109–2114 (2015).
- Liu, H. *et al.* Causes and consequences of crossing-over evidenced via a high-resolution recombination landscape of the honey bee. *Genome Biol.* **16**, 15 (2015).
- Amos, W. Heterozygosity and mutation rate: evidence for an interaction and its implications. *Bioessays* **32**, 82–90 (2010).
- Hollister, J. D., Ross-Ibarra, J. & Gaut, B. S. Indel-associated mutation rate varies with mating system in flowering plants. *Mol. Biol. Evol.* **27**, 409–416 (2010).
- Amos, W. Even small SNP clusters are non-randomly distributed: is this evidence of mutational non-independence? *Proc. R. Soc. Lond. B* **277**, 1443–1449 (2010).
- Tian, D. *et al.* Single-nucleotide mutation rate increases close to insertions/deletions in eukaryotes. *Nature* **455**, 105–108 (2008).
- Pal, C., Maciá, M. D., Oliver, A., Schachar, I. & Buckling, A. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* **450**, 1079–1081 (2007).
- Cox, E. C. On the organization of higher chromosomes. *Nature New Biol.* **239**, 133–134 (1972).
- Shee, C., Gibson, J. L. & Rosenberg, S. M. Two mechanisms produce mutation hotspots at DNA breaks in *Escherichia coli*. *Cell Rep.* **2**, 714–721 (2012).

25. Pál, C. & Hurst, L. D. Evidence for co-evolution of gene order and recombination rate. *Nature Genet.* **33**, 392–395 (2003).
26. Gladyshev, E. & Kleckner, N. Direct recognition of homology between double helices of DNA in *Neurospora crassa*. *Nature Commun.* **5**, 3509 (2014).
27. Boateng, K. A., Bellani, M. A., Gregoret, I. V., Pratto, F. & Camerini-Otero, R. D. Homologous pairing preceding SPO11-mediated double-strand breaks in mice. *Dev. Cell* **24**, 196–205 (2013).
28. Malkova, A. & Haber, J. E. Mutations arising during repair of chromosome breaks. *Annu. Rev. Genet.* **46**, 455–473 (2012).
29. Ichijima, Y. *et al.* MDC1 directs chromosome-wide silencing of the sex chromosomes in male germ cells. *Genes Dev.* **25**, 959–971 (2011).
30. Cao, J. *et al.* Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nature Genet.* **43**, 956–963 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the National Natural Science Foundation of China (91331205, 91231102 and 31170210), Program for Changjiang Scholars and Innovative Research Team (IRT_14R27) and Jiangsu Collaborative Innovation Center for Modern Crop Production to D.T., S.H. and J.-Q.C., respectively.

Author Contributions D.T., L.D.H., S.Y. and J.-Q.C. designed the experiments. S.Y., L.W., J.H., X.Z., L.D.H. and Y.Y. performed the experiments and analysed the data. L.D.H. and D.T. wrote the paper.

Author Information All Illumina reads have been deposited in the BioProject database under accession numbers PRJNA243018, PRJNA252997, PRJNA178613 and PRJNA232554. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.T. (dtian@nju.edu.cn) or L.D.H. (bssldh@bath.ac.uk).

METHODS

Materials and sequencing. We selected two purebred parents in both *Arabidopsis* (strains Col and Ler) and rice (strains 9311 and PA64s) to cross to generate intraspecific F_1 heterozygotes. Col and Ler were female and male parents, respectively. In rice, maternal PA64s and paternal 9311 were crossed to generate their F_1 progeny, the super-hybrid rice LYP9 (ref. 31). A single heterozygous F_1 seed in each species was used to generate F_2 progeny. In *Arabidopsis*, two F_2 plants (lines c52 and c64) were used to generate F_3 and F_4 plants by self-crossing. A total of 67 *Arabidopsis* and 12 rice F_2 plants and 32 *Arabidopsis* F_4 plants were randomly selected for sequencing (Fig. 1 and Extended Data Fig. 1a). In addition, the self-crossed homozygous progeny from each pure parent ($P_0 \rightarrow P_1$) were sequenced (seventeen Col, nine Ler, three 9311 and three PA64s). Finally, one each of the four parents and one F_1 (in rice) were also sequenced, making a total of 148 plants. Of these, the F_2 and F_4 plants experienced one and three meioses since F_1 , respectively (Fig. 1 and Supplementary Table 1). Col and Ler seeds were gifts from J. Bergelson. *Oryza sativa* cultivars PA64s and 9311 were obtained from C. Wang.

Two DNA samples were extracted separately from two leaves using the cetyltrimethyl ammonium bromide (CTAB) method and sequenced independently for each of *Arabidopsis* parents, their 33 F_2 progeny and all rice plants at BGI-Shenzhen. One DNA sample for the other 34 *Arabidopsis* plants was sequenced. For all, paired-end sequencing libraries with insert size of 500 bp were constructed for each DNA sample according to the manufacturer's instructions. Then, 2×100 bp paired-end reads were generated on Illumina HiSeq 2000.

For the analysis in honey bees (*Apis mellifera ligustica* Spinola), 3 queens and 43 drones were collected from 3 colonies in a bee farm (details described previously¹⁷).

The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Reads mapping and identification of candidate mutations. The Col genome (TAIR10) was downloaded from the TAIR website (http://ftp.Arabidopsis.org/home/tair/Sequences/whole_chromosomes). The assembly Ler scaffolds, SNPs and indels were downloaded from 1001 Genomes (<http://1001genomes.org/projects/assemblies.html>). The repeat and non-repeat sequences in the genome were grouped by both annotated transposable elements, RepeatMasker regions for *Arabidopsis* (<http://www.repeatmasker.org>) and homologous fragments (identity >70%; alignment length >200 bp). Raw reads were cleaned by trimming adaptor sequences and removing reads that contain more than 50% low-quality bases (quality value ≤ 5). All cleaned reads were mapped to the TAIR10 reference genome after trimming and removing low-quality bases by using the BWA-MEM (version 0.7.10) algorithm, which shows better performance than several other read aligners to date while mapping 100 bp sequences³². The mapping results were processed using Picard MarkDuplicates to remove over-sequenced DNA molecules. Mapping artefacts introduced while aligning reads on the edges of indels were removed using the GATK package^{33,34}.

After that, the HaplotypeCaller in the GATK package, which incorporates local reassembly of haplotypes, was employed to call SNPs and indels. This heavily tested protocol, used in the 1000 Genomes Project, was chosen as it provides the best reduction in false positives³⁵. We joint genotyped the relevant cohort with all *Arabidopsis* or rice samples and filtered out those sample-specific loci as the initial candidate sets. In these sets, the regions without reads in the parent samples or >8 other samples were excluded.

To ensure the accuracy of calling the *de novo* mutations, numerous stringent strategies were employed (Extended Data Fig. 1a): (1) in each sample, the candidate 'mutation' cannot be called in other non-descendent samples; (2) the candidate mutation must be called in at least 5 reads and must include both the forward and reverse reads with high variant quality score (≥ 30 for indel and ≥ 50 for SNP); (3) owing to alignment difficulties in the vicinity of indels, those base mutations located around indels (<10 bp each side) between the two parental genomes were removed; (4) the called indels that have an ≤ 20 bp interval between them were discarded. All alignments were manually inspected in Integrative Genomics Viewer (IGV)³⁶. For size distribution of indels see Supplementary Table 8.

Estimation of the possible false positives. The initial filtering may retain a number of false positives due to sequencing, mapping or genotyping errors. We employ a strategy that minimizes the false positive rate, but by necessity probably generates a higher false negative rate. While most of the errors are position-dependent, the mapping errors are less likely to show up in only a single individual in multi-independent samples³⁷. Therefore, for any given focal mutation in a focal individual, we examined the reads from the same location in all other members of the cohort and removed those 'mutations' where some reads carry the mutation allele in non-focal individuals (excepting descendants). This method becomes especially efficient with increasing sample size. For example, with our >100 samples in *Arabidopsis* derived from a single source, all individuals should share the same error rate at the same position. Hence a mutation called in one and only one F_2 is likely to be real. This method is similar to the consensus approach⁷, which is

ideal with a large number of samples and is robust against sequencing or alignment errors presenting a very low false positive rate⁷.

In addition, we extracted all reads containing candidate mutation loci, and aligned them to the reference sequence in this region using Clustalw v.2.0 (ref. 38). All alignments of each mutation-associated region were manually inspected by IGV³⁶ to minimize the risk of alignment artefacts and mapping errors. If a region has no companion in the reference genome it is ignored, possibly causing false negatives.

Furthermore, in theory, all of these mutations detected in P_1 and F_2 samples should be heterozygous (the probability that the same mutation occurs in the same position of the genome in two independent meioses is negligible). As expected, only 17 (5.6%) out of the 304 mutations were reported as 'homozygous'. The residual homozygosity might be caused by biased library construction. In fact, as expected, most (15) of them have a total depth less than or around half of the sequencing depth. These mutations were all verified by PCR as present in the F_2 but absent in the parent (P_0).

Next, a true mutation must be heritable and segregating in its progeny but any sequencing error should not be. As expected we detected about half of the mutations called in the F_2 generation in their offspring (21 F_3 progeny were randomly sampled from 41 F_2 samples with seeds and 32 F_4 samples in Fig. 1). In addition, we exclude the possibility of these mutations being present in their parents by PCR amplification and Sanger sequencing.

Last, the errors could come from a time before the sequencing owing to somatic mutation, library construction or DNA amplification at an earlier stage. These cases can be estimated by independent DNA extraction and sequencing for the same sample. The 51 plant individuals, each of which has been sequenced twice using DNA samples extracted separately from two leaves, provides an opportunity to test for possible false positives caused before sequencing. On the basis of those sequences, we found that all the mutations detected are present in both of the independent sequencing libraries.

Estimation of possible false negatives. While the next-generation sequencing mapping-based method has good accuracy and a low false positive rate in detecting candidate mutations when applied with stringent filtering^{6,8,9}, the false negative rate remains difficult to estimate accurately, but given our stringency it is likely to be considerably higher than our false positive rate. Some false negatives also appear because of technological limitations. For example, that we observe ~5.8% of F_2 mutations as being 'homozygous', suggests that we could be missing mutations because they are appearing in the unsequenced component.

We took several approaches to attempt to estimate the false negative rate. In the first we applied the method of simulating mutations described previously³⁹. In brief, 1,000 synthetic mutations were simulated by modifying sequencing reads for randomly selected sites in 20 *Arabidopsis* F_2 plants. Then, we realigned and analysed the modified data using the same procedures as for the real data. Among these 1,000 synthetic mutations, 897 were considered as callable sites according to the criteria³⁹. Finally, 880 out of the 897 sites (~98.1%) were directly identified as mutations using our pipeline, suggesting a low frequency of false negatives among callable sites (1.9%). This does not, however, address the problem of mutations missing when the sequence is missing. Indeed, 12% of sites (120 of 1,000) are missing.

A more direct way to estimate the false negative rate is to search for mutations found in more than one F_4 progeny, with these F_4 plants being derived from different F_3 plants but the same F_2 . Such shared mutations most likely were in the F_2 but missed. We can then ask how many we missed in the F_2 . In total, we identified 11 shared mutations, of which 10 were correctly detected in F_2 ancestors. PCR and Sanger sequencing confirms that the newly identified mutation is really present, but not originally called, in the F_2 . This suggests a 9.1% ($1/(10+1) = 0.091$) lower estimation of mutation rate due to false negatives.

The relationship between the divergence and diversity. The whole-genome alignments between *A. lyrata* and *A. thaliana* Col were downloaded from the VISTA database⁴⁰. Only alignments over 5,000 bp were taken for further analysis. Non-unique alignments were discarded. First, the potential substitutions between *A. lyrata* and *A. thaliana* Col were called. To this end, if the site of substitution was detected as a polymorphic site in the 80 *A. thaliana* ecotypes³⁰, it was removed (masked) before estimating the divergence between *A. thaliana* and *A. lyrata*. Thus, only the remaining substitutions, which we presume to be fixed within the population of *A. thaliana*, were used to calculate the divergence between *A. thaliana* and *A. lyrata*. This was done to remove circularity in the divergence-diversity analysis. Only the single base changes at intergenic, intron and fourfold degenerate sites were used to estimate the divergence and diversity.

The intraspecific diversity in any pairwise between-strain comparison was defined as the proportion of relevant sites that are polymorphic per window (that is, polymorphism density). The average diversity in the above regions among the 80 *A. thaliana* ecotypes was calculated in their corresponding regions of the alignments between *A. lyrata* and *A. thaliana* Col. The between-ecotype diversity

was then defined as the mean pairwise diversity comparing each of the 80 *A. thaliana* ecotypes to each other. The divergence between *A. thaliana* and *A. lyrata* was estimated using baseml with TN93 substitution model implemented in PAML⁴¹.

Calculation of the mutation rate in the meiosis of $F_2 \rightarrow F_3$ and $F_3 \rightarrow F_4$. In this study, 18 F_3 plants originating from 2 F_2 plants (c52 and c64) were selfed to produce 34 different F_4 plants. We define EM3 as the expectation of specific mutations in each F_3 and EM4 as the expectation of specific mutations in each F_4 . The mutations shared in F_4 plants are deduced to be the meiosis mutations of $F_2 \rightarrow F_3$ (Extended Data Fig. 1b). However, some of the meiosis mutations of $F_2 \rightarrow F_3$ have been lost due to chance segregation or have been classified as F_4 -specific mutations due to absence in other F_4 plants. Therefore, there may be an overestimate for these F_4 -specific mutations generated in the meiosis of $F_3 \rightarrow F_4$, and there may be an underestimate for these shared mutations in those F_4 plants that have been generated in the meiosis of $F_2 \rightarrow F_3$.

Specifically, one-quarter of the mutations present in the germ line before the specialization of the reproductive tissues are expected to be homozygous at the beginning of the next generation. Let μ be the estimate of new homozygous mutations per generation, and τ , the estimate of new heterozygous mutations per generation⁸. μ_3 : μ from F_2 to F_3 ; μ_4 : μ from F_2 to F_4 ; τ_3 : τ from F_2 to F_3 ; τ_4 : τ from F_3 to F_4 .

For one generation, the estimate of mutations = $N\mu + N\tau$, where N is the number of organisms in this generation.

For two generations, the estimate of mutations = $N_1\mu_1 + N_1P\tau_1 + N_2\mu_2 + N_2\tau_2$. Here P is the probability that the heterozygous mutation in the first generation is inherited by its progeny, which depends on the number of progeny.

In our study, as we have 6 F_3 plants with 1 progeny, 10 F_3 plants with 2 progeny and 2 F_3 , the formula can be changed to: all mutations observed in $F_4 = 18\mu_3 + (6P_1 + 10P_2 + 2P_3)\tau_3 + 32\mu_4 + 32\tau_4$, where P_n is the likelihood that a mutation in an F_3 with n progenies is inherited.

For a heterozygous mutation in F_3 with n progeny, the probability that no progeny genotype is a homozygous mutation (a/a) is 0.75^n , and that at least one of the progeny carry a homozygous mutation is $1 - 0.75^n$.

For all the homozygous mutations in F_4 :

$$18\mu_3 + 32\mu_4 + [6 \times (1 - 0.75) + 10 \times (1 - 0.75^2) + 2 \times (1 - 0.75^3)]\tau_3 = 19 + 3 \quad (1)$$

For a heterozygous mutation in F_3 , the probability that it is not inherited by the progeny is 0.25^n and the probability that the mutation appears as heterozygous in F_3 plants is $0.75^n - 0.25^n$. For F_3 with 1, 2 or 3 progenies, the likelihood is 0.5, 0.5 and 0.40625, respectively.

For all the heterozygous mutations in F_4 :

$$(6 \times 0.5 + 10 \times 0.5 + 2 \times 0.40625)\tau_3 + 32\tau_4 = 54 + 6 \quad (2)$$

The shared mutations in F_4 can be counted as the result of mutations in F_3 .

As shown (Extended Data Fig. 1b) for the shared heterozygous mutations in F_4 :

$$[10 \times 0.25 + 2 \times (0.5^3 + 3 \times 0.5^2 \times 0.25)]\tau_3 = 6 \quad (3)$$

$$\tau_3 = 1.92$$

According to equations (2) and (3):

$$\tau_4 = 1.346$$

According to equations (1) and (3):

$$18\mu_3 + 32\mu_4 = 8.5 \quad (4)$$

$$\mu_3 + 1.78\mu_4 = 0.472$$

$$\mu_3 \leq 0.472$$

$$\mu_4 \leq 0.266$$

If a homozygous mutation occurred in 10 F_3 plants with 2 progeny or 2 F_3 plants with 3 progeny (counted as μ_3), all of its progeny will carry homozygous mutations, which was not found in our result, so μ_3 was assumed to be 0.

$$\mu_3 = 0$$

$$\mu_4 = 0.266$$

$$EM3 = \mu_3 + \tau_3 = 1.92$$

$$EM4 = \mu_4 + \tau_4 = 1.612$$

Therefore, the mutation rates of F_2 to F_3 or F_3 to F_4 should be 1.60×10^{-8} or 1.34×10^{-8} , respectively.

Distribution of mutations and statistical analyses. To determine the distribution of mutations on chromosomes (Extended Data Fig. 3d), the *de novo* mutations were used from our sequenced 26 P_1 , 67 F_2 and 32 F_4 plants, and two published data sets^{8,9}, all of which employed the ecotypes of Col, Ler or the offspring of Col \times Ler. The recombination (crossover) data were collected from our 67 F_2 and 32 F_4 plants (Supplementary Table 9).

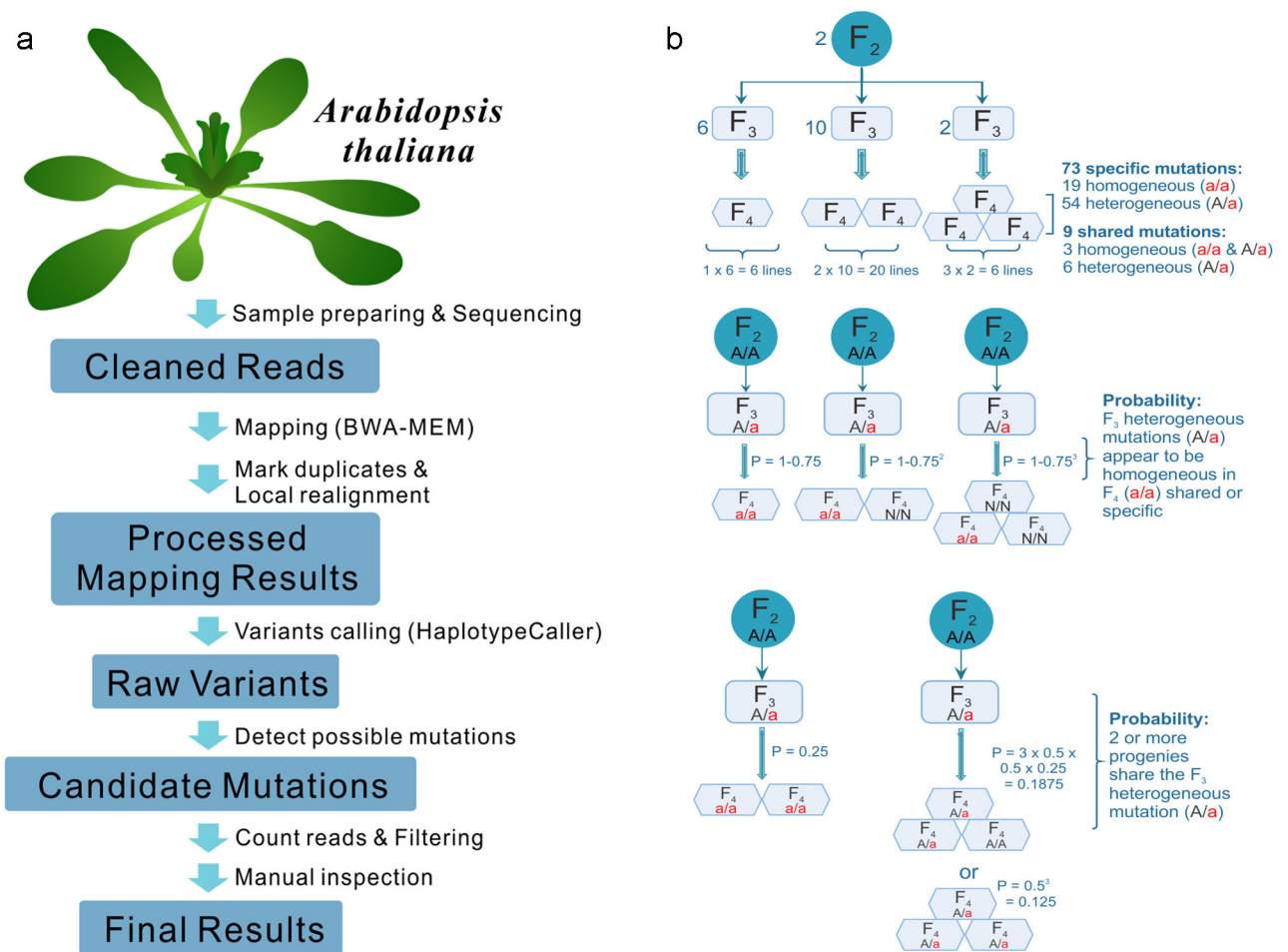
To determine whether proximity to heterozygous sites could affect the mutation rate, we calculated the distance of the new mutations to heterozygous sites. To detect whether the observed mutations tend to arise on derived versus ancestral alleles, we make use of the alignments, described earlier, between *A. thaliana* (Col) and *A. lyrata*. If the same aligned nucleotide is seen in both *A. lyrata* and *A. thaliana* (before mutation) it was presumed to reflect the ancestral state. A total of 201 mutations (158 SNPs and 43 indels) have a clear ancestral state. Of the remaining 199, 93 have no alignments, 59 are in the gaps of *A. lyrata*, 15 are ambiguous due to non-unique alignments, and 32 have a different nucleotide compared to *A. thaliana* thus preventing ancestral state determination.

To estimate the expected number of mutations in heterozygous and homozygous compartments under a null expectation that heterozygosity *per se* is not a relevant parameter, we factor in both the absolute size of both compartments and, for point mutations, the trinucleotide content. The GC content of sequence-flanking indels (35%) is almost identical to that of the genomic average (36%) so we make no nucleotide content correction for these. Given a total observed set of mutations, we calculate a mutation rate per given trinucleotide triplet, with the mutation centred with the triplet. We then, for each compartment, calculate the total number of each triplet to generate an expected number of point mutations per triplet. We then sum across all triplets to derive an expected total number of mutations in a given compartment. As an internal consistency check we calculate the sum across the two compartments, ensuring that this is the same as the observed total number of mutations. We thus have both observed and expected (allowing for nucleotide content and span length) number of mutations. For indels we just consider the proportion of all sequences in each compartment. We test for difference by chi-squared test with Yate's correction.

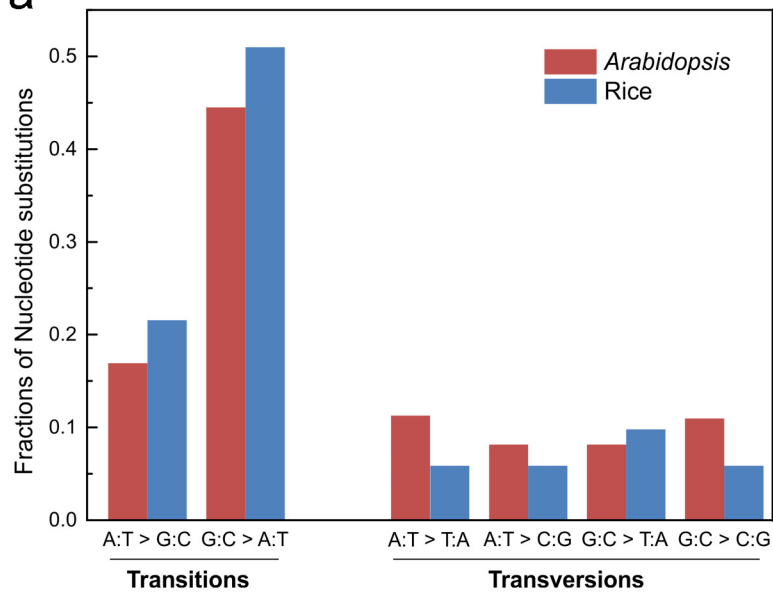
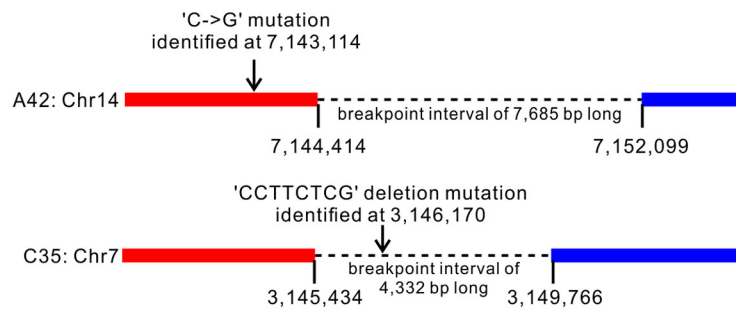
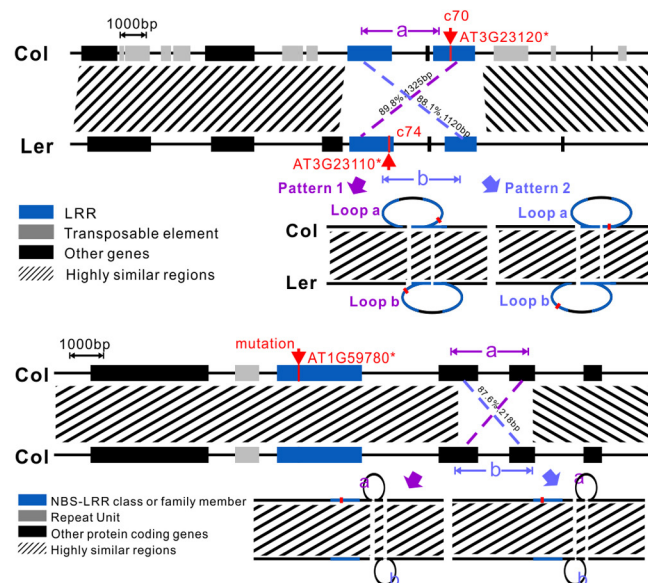
Statistics were performed in R⁴². Brunner–Munzel test was implemented in lawstat package. When P values were derived from randomization, 10,000 randomizations were employed in which the data were randomly ascribed by shuffling of class (for example, heterozygous or homozygous). The unbiased estimation of empirical P , meaning expected type I error rate, is $(n + 1)/(m + 1)$, where n is the number of observations as or more extreme than that observed in the real test reporting statistic and m is the number of randomization⁴³.

We used the prior mutation rate estimate to inform the sort of sample sizes needed, but no statistical methods were used to predetermine sample size.

- Gao, Z.-Y. *et al.* Dissecting yield-associated loci in super hybrid rice by resequencing recombinant inbred lines and improving parental genome sequences. *Proc. Natl Acad. Sci. USA* **110**, 14492–14497 (2013).
- Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <http://arxiv.org/abs/1303.3997> (2013).
- DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **43**, 491–498 (2011).
- McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
- Ghoniem, D. H., Myers, J. R., Tuttle, E. & Paciorkowski, A. R. Comparison of insertion/deletion calling algorithms on human next-generation sequencing data. *BMC Res. Notes* **7**, 864 (2014).
- Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
- Li, M. & Stoneking, M. A new approach for detecting low-level mutations in next-generation sequence data. *Genome Biol.* **13**, R34 (2012).
- Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
- Keightley, P. D., Ness, R. W., Halligan, D. L. & Haddrell, P. R. Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. *Genetics* **196**, 313–320 (2014).
- Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. & Dubchak, I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* **32**, W273–W279 (2004).
- Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
- R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2013).
- North, B. V., Curtis, D. & Sham, P. C. A note on the calculation of empirical P values from Monte Carlo procedures. *Am. J. Hum. Genet.* **72**, 498–499 (2003).

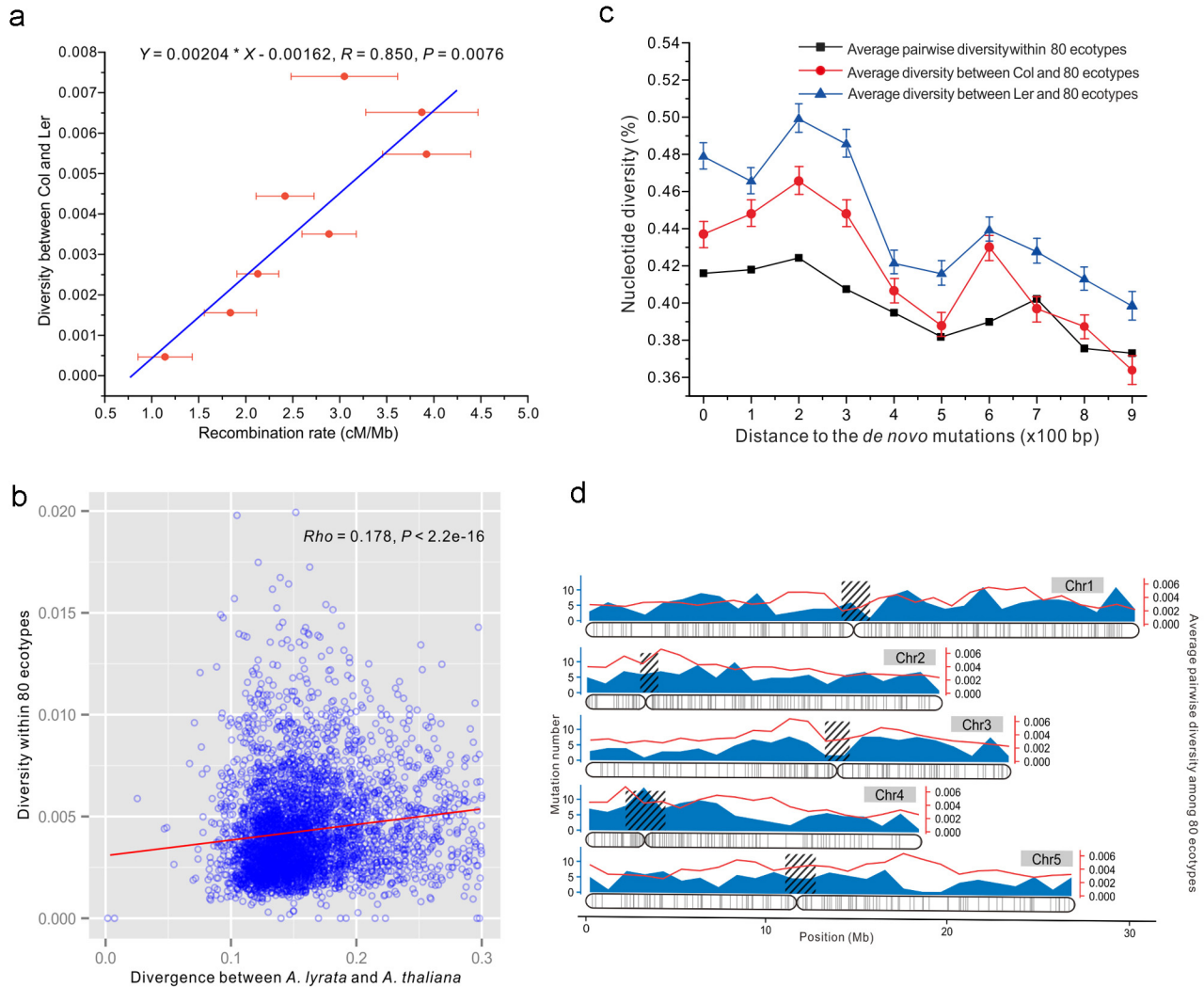


Extended Data Figure 1 | Details of materials and methods. **a**, Schematic diagram of the detection of *de novo* mutations. **b**, The calculation of the expected mutations in the meiosis of $F_2 \rightarrow F_3$ (EM3) and $F_3 \rightarrow F_4$ (EM4). For further explanation see Methods.

a**b****c**

Extended Data Figure 2 | Mutational properties. **a**, Spectra of nucleotide substitutions in *Arabidopsis* and rice. **b**, Co-occurrence of mutations and crossover break points in bees. By using the sequence data of 43 honey bee drones and their 3 corresponding queens¹⁷, a total of 27 base and 8 indel mutations were detected. Of note, 2 of 35 mutations are found in close proximity with crossover break points in the same sample (distance < 2 kb; $P = 0.0012$ with 10,000 randomizations), these being illustrated here. The crossover event is between the red and blue line with marker positions annotated. The positions of the mutations are annotated with arrows. **c**, A schematic diagram of the genomic structures and the possible pairings of two homologous chromosomes during the meiosis at two mutated *LRR-TM* genes (top) and one mutated *NBS-LRR* gene (bottom). The top panel shows the

genomic structures between Col and Ler at the loci of *AT3G23110*, the receptor-like protein 37 with a non-synonymous mutation (Chr3:8224726, T→C) at sample of c74, and *AT3G23120*, the receptor-like protein 38 with a deletion mutation (Chr3:8228194, Del:C, frameshift) at sample of c70. The bottom panel illustrates the genomic structures between two Col chromosomes at *AT1G59780* and the mutations detected in a homozygous plant of Col⁸. Red arrows represent the position of mutation; the hatched areas indicate the highly similar sequences, the other regions being highly diversified; the dotted lines indicate the paired length of the homologues at the highly identical regions. During meiosis, possible pairings between parental chromosomes are illustrated, where the loops indicate the unpaired regions.



Extended Data Figure 3 | Correlation between mutations, recombination events, diversity and divergence. **a**, The relationship between nucleotide diversity (Col versus Ler) and recombination rate. When the chromosomes were dissected into 100 kb non-overlapping windows, the diversity (polymorphism density) between Col and Ler and the recombination rates in 67 F_2 and 32 F_4 plants were calculated for each window. When sorting the windows by the diversity and dividing them into 8 equal intervals (for example, from 0 to 0.001, 0.001 to 0.002, 0.002 to 0.003, and so on), the relationships between the average diversity and recombination rate is displayed. Error bars indicate standard error of the mean. **b**, The relationship between diversity and divergence. The red line represents standard linear regression and is for illustrative purposes only. The statistic is the result of Spearman's rank correlation. **c**, Relationship between mutation and distance to polymorphic sites. The mutation data were collected from our 67 F_2 samples. Window 0 in

the x-axis is the 2×100 bp sequence surrounding the position of any given *de novo* mutation and 1–9 is 100–900 bp away from the mutation on both sides. For each window of 2×100 bp sequence, the average diversity is calculated. The black squares denote the average pairwise diversity among the published 80 *Arabidopsis* ecotypes; the red circles denote the average diversity between Col and the 80 ecotypes; the blue triangles denote the average diversity between Ler and the 80 ecotypes. Error bars indicate standard error of the mean. **d**, Distribution of the mutations on the chromosomes. The grey vertical bars in the chromosomes denote the position of all collected mutations. When the chromosomes were dissected into 1 Mb non-overlapping windows, the mutation numbers (blue shadow in the figure) were counted in each window. The red lines denote the average pairwise diversity among the published 80 *Arabidopsis* ecotypes.

Epoxyeicosatrienoic acids enhance embryonic haematopoiesis and adult marrow engraftment

Pulin Li^{1,2*}, Jamie L. Lahvic^{1*}, Vera Binder^{1,3*}, Emily K. Pugach¹, Elizabeth B. Riley¹, Owen J. Tamplin¹, Dipak Panigrahy⁴, Teresa V. Bowman¹, Francesca G. Barrett¹, Garrett C. Heffner¹, Shannon McKinney-Freeman⁵, Thorsten M. Schlaeger¹, George Q. Daley¹, Darryl C. Zeldin⁶ & Leonard I. Zon^{1,2}

Haematopoietic stem and progenitor cell (HSPC) transplant is a widely used treatment for life-threatening conditions such as leukaemia; however, the molecular mechanisms regulating HSPC engraftment of the recipient niche remain incompletely understood. Here we develop a competitive HSPC transplant method in adult zebrafish, using *in vivo* imaging as a non-invasive readout. We use this system to conduct a chemical screen, and identify epoxyeicosatrienoic acids (EETs) as a family of lipids^{1,2} that enhance HSPC engraftment. The pro-haematopoietic effects of EETs were conserved in the developing zebrafish embryo, where 11,12-EET promoted HSPC specification by activating a unique activator protein 1 (AP-1) and *runx1* transcription program autonomous to the haemogenic endothelium. This effect required the activation of the phosphatidylinositol-3-OH kinase (PI(3)K) pathway, specifically PI(3)K γ . In adult HSPCs, 11,12-EET induced transcriptional programs, including AP-1 activation, which modulate several cellular processes, such as migration, to promote engraftment. Furthermore, we demonstrate that the EET effects on enhancing HSPC homing and engraftment are conserved in mammals. Our study establishes a new method to explore the molecular mechanisms of HSPC engraftment, and discovers a previously unrecognized, evolutionarily conserved pathway regulating multiple haematopoietic generation and regeneration processes. EETs may have clinical application in marrow or cord blood transplantation.

To our knowledge, a screen-based forward-genetic approach to understand transplantation biology has never been attempted. In an effort to quantify HSPC activity, we developed a competitive transplantation system in a transparent mutant zebrafish, *casper*³, which allows direct visualization of engraftment *in vivo*. We co-injected whole kidney marrow (WKM) cells from two ubiquitous GFP and DsRed2 transgenic donors into *casper* (Fig. 1a), and calculated relative engraftment as the ratio of GFP/DsRed2 intensity (G/R) within the same kidney region (Fig. 1b). We validated the quantitative potential of this imaging-based approach by comparing with flow cytometry-based analysis of WKM from the same recipient (Fig. 1c). The assay was also sensitive to changes in the relative number of green-to-red donor cells (Fig. 1d). Additionally, our system successfully detected the effects of two known chemical modulators of HSPC engraftment: dmPGE₂ (16,16-dimethyl-prostaglandin E₂), a stabilized derivative of PGE₂ (ref. 4), and BIO (6-bromoindirubin-3'-oxime), a GSK-3 β inhibitor⁵. We used our assay to screen 480 compounds with known bioactivities, which had been selected to cover diverse signalling pathways (Extended Data Fig. 1a). Ten compounds significantly increased the G/R ratio reproducibly, including PGE₂ and Ro 20-1724, which activates the cAMP pathway downstream of PGE₂ (refs 4 and 5). The other

hits target pathways that previously have not been linked to HSPC engraftment, including 11,12-EET and 14,15-EET (Fig. 1e). These are arachidonic-acid-derived eicosanoids that are synthesized through the cytochrome P450 epoxygenase pathway^{1,2} (Extended Data Fig. 1b).

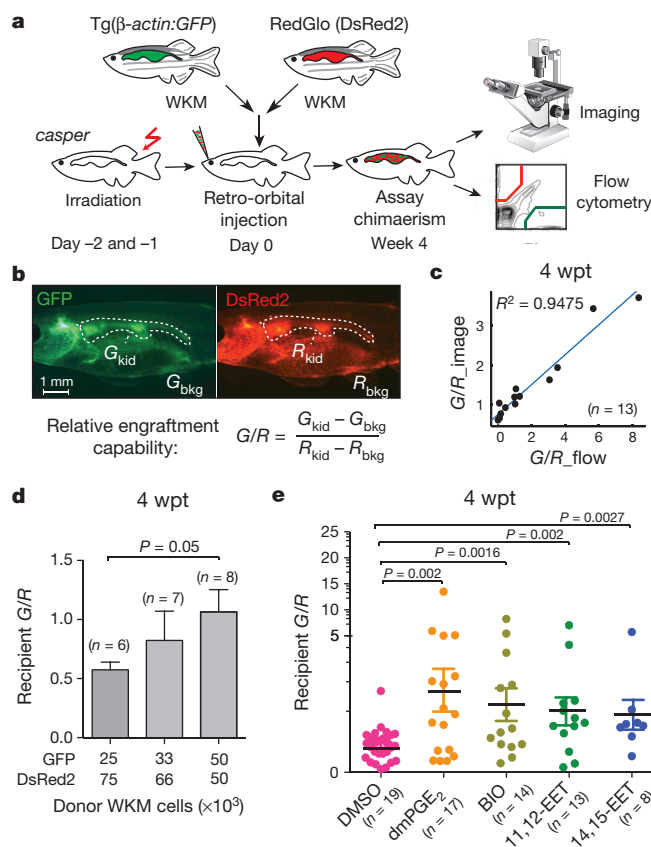


Figure 1 | Zebrafish whole kidney marrow competitive transplantation-based chemical screen identifies EETs as enhancers of marrow engraftment.

a, Schematic of zebrafish whole kidney marrow (WKM) competitive transplantation. **b**, Calculation of relative engraftment capability (G/R). White dashed line denotes kidney. G_{kid}/R_{kid} , kidney fluorescence intensity; G_{bkg}/R_{bkg} , background fluorescence intensity. **c**, The G/R ratios from imaging linearly correlated with flow cytometry analysis of the same recipients (linear regression). wpt, weeks post-transplant. **d**, Serial dilution competitive transplantation with varying donor GFP/DsRed2 ratios. **e**, Four-hour transient chemical treatment increased WKM engraftment. 11,12- and 14,15-EET, 0.5 μ M. Unpaired two-tailed *t*-test; mean and s.e.m. (d, e).

¹Stem Cell Program and Division of Haematology/Oncology, Boston Children's Hospital and Dana-Farber Cancer Institute, Howard Hughes Medical Institute, Harvard Stem Cell Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Chemical Biology Program, Harvard University, Cambridge, Massachusetts 02138, USA. ³Department of Hematology and Oncology, Dr. von Hauner Children's Hospital, Ludwig-Maximilians University, 80337 Munich, Germany. ⁴Center for Vascular Biology Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁵Department of Haematology, St Jude Children's Research Hospital, Memphis, Tennessee 38105-3678, USA. ⁶Division of Intramural Research, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina 27709, USA.

*These authors contributed equally to this work.

A gene expression study previously reported mouse *Cyp2j6*, a cytochrome P450 epoxidegenase, as one of the 93 genes enriched in long-term haematopoietic stem cells⁶.

Despite years of research on the potent effects of EETs in numerous physiological processes^{7–9,22}, knowledge about their direct target(s) and downstream pathway(s) is still very limited. To tackle this problem, a robust system allowing easy genetic perturbation is crucial. As adult regeneration often reactivates pathways important for development, we decided to probe the effects of EETs on haematopoiesis during embryo development. Analogous to mammalian development, zebrafish HSPCs form from a *flk1*⁺ population, named haemogenic endothelium, at 24 hours post fertilization (hpf), and become *runx1*⁺ at 36 hpf in the evolutionarily conserved aorta–gonad–mesonephros (AGM) region^{10–12}. HSPCs enter the circulation after they emerge from the AGM^{11–13}, and seed the caudal haematopoietic tissue (CHT), a secondary haematopoietic site equivalent to the mammalian fetal liver^{14,15} (Fig. 2a). The 11,12-EET treatment between 24 and 36 hpf strongly increased the HSPC marker *runx1* in the AGM, and surprisingly induced *runx1* in a non-haematopoietic region of the tail mesenchyme, where *runx1* is not normally expressed (Fig. 2b). This indicates 11,12-EET might be inducing a conserved transcriptional program. We confirmed this AGM phenotype with *in vivo* time-lapse imaging of HSPC birth from the haemogenic endothelium. Tg(*CD41:GFP; flk1:DsRed2*) embryos treated with 11,12-EET starting at 24 hpf showed a significant increase in the number of double-positive HSPCs in the AGM from 30 to 46 hpf (Fig. 2c, d). Single-cell analysis showed that this change is mainly due to a significant increase in the frequency of HSPCs directly specified from the haemogenic endothelium, while no increase in the rate of cell division or AGM

retention was observed (Extended Data Fig. 2). The additional HSPCs produced after 11,12-EET treatment successfully homed to their next niche, resulting in increased numbers of HSPCs in the CHT, which was verified by *in situ* hybridization for the HSPC marker *cmyb* (Fig. 2e and Extended Data Fig. 3). Time-lapse imaging of Tg(*Runx1+23:GFP*) zebrafish showed that 11,12-EET treatment between 24 and 48 hpf increased the rate of arrival of GFP⁺ HSPCs to the CHT (Fig. 2f and Supplementary Videos 1 and 2), presumably owing to enhanced HSPC specification in the AGM.

To dissect the molecular mechanism leading to *runx1* induction further, we performed microarray analysis on 11,12-EET-treated 36-hpf embryos (Supplementary Table 3). The upregulation of multiple activator protein 1 (AP-1) family transcription factors, including *fosl2*, and duplicated orthologues of human JUNB, *junb* and *junbl*, were among the most prominent changes. Whole-mount *in situ* hybridization confirmed the induction both in the AGM and the non-haematopoietic region of the tail mesenchyme (Fig. 3d, top two rows). AP-1 messenger RNA transcripts were detectable within 1 h of 11,12-EET treatment and insensitive to the protein translation inhibitor cycloheximide (Extended Data Fig. 4a, b), indicating that AP-1 members are immediate targets of EET signalling. By contrast, *runx1* induction required at least 4 h of 11,12-EET treatment and was completely blocked by cycloheximide (Extended Data Fig. 4c). Therefore, we proposed that EET-induced AP-1 expression is necessary for increasing *runx1* transcription.

To test this hypothesis genetically, we globally knocked down AP-1 with anti-sense morpholinos targeting *junb* and *junbl*, which blocked *runx1* expression without affecting endothelial cells of the AGM (Extended Data Fig. 5), suggesting that AP-1 might be required for HSPC specification from haemogenic endothelium. To test whether AP-1 function is autonomous to the haemogenic endothelium, we delivered a dominant-negative form of JunB protein (dnJUNB) specifically to the *flk1*⁺ endothelial cells, before the induction of *runx1*, to functionally inhibit all AP-1 activity. Although *flk1:dnJUNB* did not significantly reduce the expression of *runx1* in DMSO-treated embryos, it suppressed the EET-induced increase of *runx1* in the AGM (Fig. 3a, b). Combined with the gene expression data, these genetic analyses showed that 11,12-EET activates an AP-1 and *runx1* transcriptional cascade of cell-fate specification autonomous to the haemogenic endothelium.

In an effort to define downstream signalling events for 11,12-EET, we performed a chemical suppressor screen in zebrafish embryos by examining the capability of various chemicals to suppress the 11,12-EET-induced AP-1 and *runx1* gene signature (Fig. 3c). Several PI(3)K inhibitors completely blocked the signature without detrimental effects to overall embryonic development (Fig. 3d, e and Extended Data Fig. 6a). To interrogate specific PI(3)K catalytic subunits, we assayed subunit-specific chemical inhibitors and morpholinos targeting individual class I PI(3)K subunits. Among α -, β -, γ - and δ -subunits of PI(3)K, only PI(3)K γ loss of function specifically abrogated the *runx1* induction in the AGM and tail non-haematopoietic tissue (Extended Data Fig. 6b, c). Furthermore, 11,12-EET enhanced PI(3)K activity in immortalized human umbilical vein endothelial cells, assayed by Akt phosphorylation (data not shown). No such increase was seen in human umbilical cord blood CD34⁺ HSPCs, although EET-induced gene expression changes could be partially blocked in these cells by co-treatment with PI(3)K inhibitors. This indicates PI(3)K functions either directly downstream of 11,12-EET or as a parallel pathway, depending on the cellular context. In either case, PI(3)K activity is required for inducing the AP-1 and *runx1* transcription cascade in the AGM.

To understand how 11,12-EET treatment leads to increased engraftment in already-specified HSPCs, we performed RNA-sequencing in human umbilical cord blood CD34⁺ HSPCs and a human myeloid cell line (U937), and used Ingenuity Pathway Analysis (IPA) to decipher the biological pathways regulated by 11,12-EET in both cell types

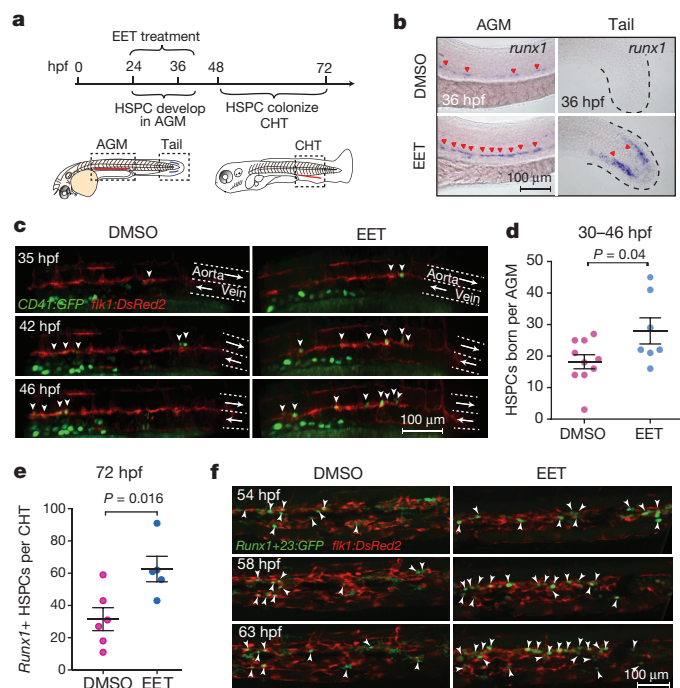


Figure 2 | 11,12-EET enhances HSPC specification in the zebrafish embryo AGM. **a**, Schematic of HSPC development in zebrafish embryos. **b**, Representative images of whole-mount *in situ* hybridization showing 11,12-EET (24–36 hpf treatment) induced HSPC marker *runx1* in the AGM and a tail non-haematopoietic tissue (>8 independent experiments, $n > 100$). **c**, **d**, 11,12-EET (24–46 hpf) enhanced *CD41:GFP/flk1:DsRed2* double-positive HSPCs (white arrowheads) emerging in the AGM. Arrows indicate blood flow. **e**, **f**, Same treatment increased the number of HSPCs in the CHT. **e**, mCherry⁺ HSPCs quantified in the Tg(*Runx1+23:mCherry*) CHT. **f**, Representative montage images of *Runx1+23:GFP* HSPCs (white arrowheads) engrafting CHT. *flk1:DsRed2*, endothelial cells. Unpaired two-tailed *t*-test, mean and s.e.m. (**d**, **e**).

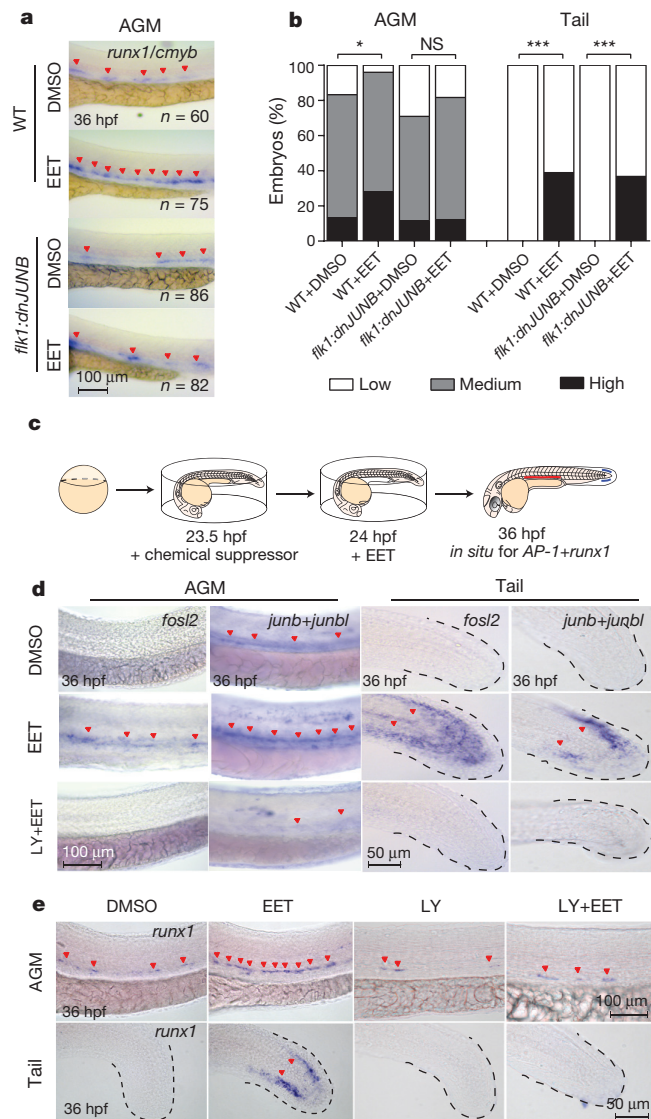


Figure 3 | 11,12-EET induces a PI(3)K-dependent AP-1/*runx1* transcriptional program to increase HSPC specification. **a**, **b**, Stable *flk1:dnJUNB*-2A-GFP expression blocking AP-1 function suppressed 11,12-EET-enhanced HSPCs in the AGM. Representative images of *runx1* and *cmyb* in situ hybridization (**a**) and quantification (**b**) after 11,12-EET treatment (24–36 hpf). Embryos scored as high, medium or low *runx1* and *cmyb*, summed across 4 experiments. **P* = 0.01, ****P* < 0.0001, Chi-square. NS, not significant; WT, wild-type. **c**, Schematic of chemical screen for EET signalling pathway suppressors. **d**, **e**, 11,12-EET induced AP-1 family transcription factors (*fosl2*, *junb* and *junbl*) (**d**) and *runx1* (**e**), suppressed by cotreatment with the PI(3)K inhibitor LY294002 (LY), in the AGM and tail (**d**, **e**) (three independent experiments, *n* > 40). Same images from Fig. 2b were used as staining controls (**e**).

(Extended Data Fig. 7 and Supplementary Table 4). Cell-to-cell signalling and cellular movement networks topped the list of activated biological pathways, including the AP-1 members, which have been shown to modulate cell migration in many cell types^{16,17}. AP-1 thus seems to be a common target of EET signalling, which leads to the induction of *runx1* in the haemogenic endothelium (Fig. 3), and probably supports cell migration and cell–cell signalling of already-specified haematopoietic cells. By contrast, *RUNX1* is not upregulated in already-specified HSPCs, which is consistent with previous studies showing that *Runx1* is dispensable for HSPCs to engraft later haematopoietic sites¹⁸. Several cytokines, such as *CXCL8*, *OSM* and *CCL2*, were also upregulated and involved in the cell migration network.

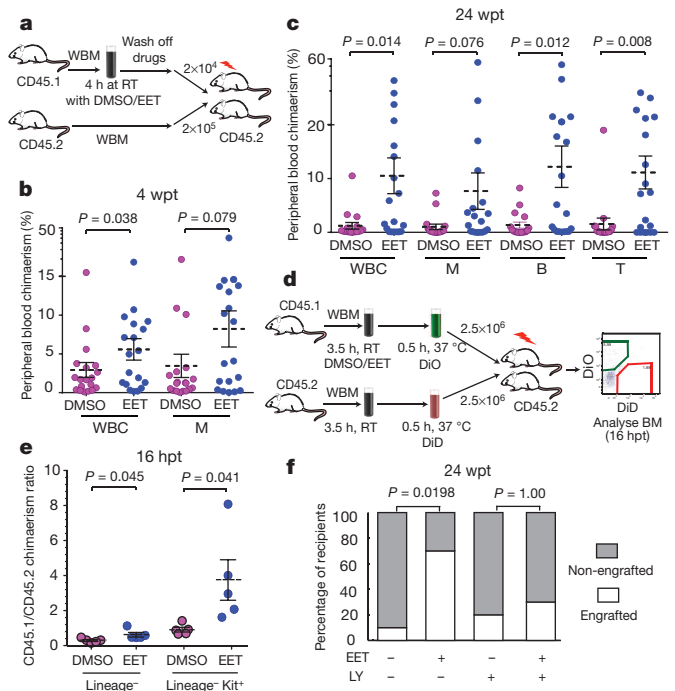


Figure 4 | 11,12-EET enhances HSPC engraftment and homing in mammals. **a**, Schematic of mouse WBM competitive transplantation. RT, room temperature. **b**, **c**, Four hours of 11,12-EET treatment promoted short-term WBM engraftment at 4 wpt (**b**) and long-term multilineage engraftment at 24 wpt (**c**). B, B cells; M, myeloid cells; T, T cells; WBC, white blood cells. Two independent experiments combined, *n* = 20 total. **d**, Schematic of WBM competitive homing assay. DiD and DiO denote cell-labelling solutions. **e**, 11,12-EET increased homing efficiency of Lin⁻ cells and Lin⁻ Kit⁺ HSPCs (*n* = 5). **f**, PI(3)K activation is required for EET-enhanced mouse WBM engraftment (*n* = 10). LY, 10 μM LY294002. Recipients characterized as engrafted or non-engrafted based on peripheral blood WBC chimaerism, two-tailed Fisher's exact test (**b**, **f**); unpaired two-tailed *t*-test (**c**, **e**), mean and s.e.m.

These data show that besides promoting HSPC specification from the haemogenic endothelium, 11,12-EET can also directly induce gene expression programs beneficial for engraftment in already-specified HSPCs. Similarly, 11,12-EET treatment of zebrafish embryos after 48 hpf, when AGM HSPC production has already completed, leads to increased HSPCs in the CHT in a PI(3)Kγ-dependent manner, without affecting cell apoptosis or proliferation (Extended Data Fig. 8). Our data strongly suggest that 11,12-EET modulates cell migration and cell–cell interaction during HSPC engraftment.

To test the evolutionary conservation of EET-induced haematopoietic phenotypes, we examined the effect of 11,12-EET on HSPC engraftment in mammalian bone marrow competitive transplantation. Consistently, 11,12-EET promoted greater short-term chimaerism by 4 weeks post-transplant compared to control-treated cells (Fig. 4a, b). Even up to 24 weeks, EET-treated marrow maintained greater multi-lineage contribution (Fig. 4c). Enhanced short- and long-term engraftment suggests that 11,12-EET may affect both stem and progenitor cells, perhaps by establishing a competitive advantage at the early stage of engraftment. In a whole-bone-marrow (WBM) homing assay, we found 11,12-EET promoted the initial seeding of progenitor cells in the bone marrow (Fig. 4d, e). The early effect could be due to an enhanced cell migration and cell–cell signalling program, since assaying cell proliferation or apoptosis in whole marrow immediately after 11,12-EET treatment did not show significant changes (Extended Data Fig. 9). However, this does not exclude the possibility of a later onset of anti-apoptotic effects on transplantation. Finally we found transient inhibition of PI(3)K partially blocked EET-induced enhancement of long-term, multi-lineage engraftment after mouse

bone marrow transplant (Fig. 4f). Thus, the EET effect on enhancing HSPC engraftment is evolutionarily conserved in fish and mammals.

Our unbiased chemical genetic studies establish a new eicosanoid pathway for haematopoiesis, which increases HSPC specification in the AGM by inducing AP-1 and *runx1*, and also enhances HSPC engraftment by modulating several biological pathways, such as migration and cell–cell signalling. Previous work in our laboratory discovered a different eicosanoid, PGE₂, could also enhance marrow engraftment^{4,5}. Both PGE₂ and EETs are arachidonic-acid-derived eicosanoids that are locally produced near wounds, and may facilitate progenitor recruitment, engraftment and proliferation. Despite their common origin, the underlying molecular signalling mechanisms and activities of PGE₂ and EETs are different (Supplementary Table 5). Although the direct receptor for EETs is unknown, several studies have provided biochemical evidence that EETs bind to a G-protein-coupled receptor (GPCR)^{19,20}. GPCRs signal through various G α subunits²¹. Previously, we showed that PGE₂ signals through the cAMP-dependent G α s-coupled PGE₂ receptor for its pro-haematopoietic effects⁵. Using chemical inhibition and genetic loss-of-function approaches, we screened all families of zebrafish G α subunits. Notably, we found that *gna12* and *gna13* are specifically required for EET-induced AP-1 and *runx1* expression (Extended Data Fig. 10). Inhibiting G α s did not suppress the EET phenotypes, indicating that EETs and PGE₂ have different signalling mechanisms.

During marrow transplantation, the achieved chimaerism over time is critical, and the time to adequate neutrophil engraftment is an important milestone for treatment success. In addition to improving long-term repopulation, EETs seem to have a prominent effect on progenitor engraftment, as shown by increased chimaerism early after transplantation. Our studies highlight the importance of lipid mediators in regulating HSPC engraftment, and the manipulation of these pathways could have clinical impact for patients undergoing transplantation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 31 October 2012; accepted 11 May 2015.

- Spector, A. A. & Kim, H.-Y. Y. Cytochrome P450 epoxygenase pathway of polyunsaturated fatty acid metabolism. *Biochim. Biophys. Acta* **1851**, 356–365 (2015).
- Node, K. *et al.* Anti-inflammatory properties of cytochrome P450 epoxygenase-derived eicosanoids. *Science* **285**, 1276–1279 (1999).
- White, R. M. *et al.* Transparent adult zebrafish as a tool for *in vivo* transplantation analysis. *Cell Stem Cell* **2**, 183–189 (2008).
- North, T. E. *et al.* Prostaglandin E2 regulates vertebrate haematopoietic stem cell homeostasis. *Nature* **447**, 1007–1011 (2007).
- Goessling, W. *et al.* Genetic interaction of PGE2 and Wnt signaling regulates developmental specification of stem cells and regeneration. *Cell* **136**, 1136–1147 (2009).
- Forsberg, E. C. *et al.* Molecular signatures of quiescent, mobilized and leukemia-initiating hematopoietic stem cells. *PLoS ONE* **5**, e8785 (2010).
- Panigrahy, D., Greene, E. R., Pozzi, A., Wang, D. W. & Zeldin, D. C. EET signaling in cancer. *Cancer Metastasis Rev.* **30**, 525–540 (2011).
- Pfister, S. L., Gauthier, K. M. & Campbell, W. B. Vascular pharmacology of epoxyeicosatrienoic acids. *Adv. Pharmacol.* **60**, 27–59 (2010).
- Wang, Y. *et al.* Arachidonic acid epoxygenase metabolites stimulate endothelial cell growth and angiogenesis via mitogen-activated protein kinase and phosphatidylinositol 3-kinase/Akt signaling pathways. *J. Pharmacol. Exp. Ther.* **314**, 522–532 (2005).
- Lam, E. Y., Hall, C. J., Crosier, P. S., Crosier, K. E. & Flores, M. V. Live imaging of Runx1 expression in the dorsal aorta tracks the emergence of blood progenitors from endothelial cells. *Blood* **116**, 909–914 (2010).
- Bertrand, J. Y. *et al.* Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature* **464**, 108–111 (2010).
- Kissa, K. & Herbomel, P. Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature* **464**, 112–115 (2010).
- Boisset, J. C. *et al.* *In vivo* imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature* **464**, 116–120 (2010).
- Murayama, E. *et al.* Tracing hematopoietic precursor migration to successive hematopoietic organs during zebrafish development. *Immunity* **25**, 963–975 (2006).
- Tamplin, O. J. *et al.* Hematopoietic stem cell arrival triggers dynamic remodeling of the perivascular niche. *Cell* **160**, 241–252 (2015).
- Renaud, S. J., Kubota, K., Rumi, M. A. & Soares, M. J. The FOS transcription factor family differentially controls trophoblast migration and invasion. *J. Biol. Chem.* **289**, 5025–5039 (2014).
- Gilan, O. *et al.* PR55 α -containing protein phosphatase 2A complexes promote cancer cell migration and invasion through regulation of AP-1 transcriptional activity. *Oncogene* **34**, 1333–1339 (2015).
- Chen, M. J., Yokomizo, T., Zeigler, B. M., Dzierzak, E. & Speck, N. A. Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature* **457**, 887–891 (2009).
- Chen, Y., Falck, J. R., Manthali, V. L., Jat, J. L. & Campbell, W. B. 20-Iodo-14,15-epoxyeicosa-8(Z)-enoyl-3-azidophenylsulfonamide: photoaffinity labeling of a 14,15-epoxyeicosatrienoic acid receptor. *Biochemistry* **50**, 3840–3848 (2011).
- Yang, W. *et al.* Characterization of epoxyeicosatrienoic acid binding site in U937 membranes using a novel radiolabeled agonist, 20-125I-14,15-epoxyeicosa-8(Z)-enoic acid. *J. Pharmacol. Exp. Ther.* **324**, 1019–1027 (2008).
- Lappano, R. & Maggiolini, M. G protein-coupled receptors: novel targets for drug discovery in cancer. *Nature Rev. Drug Discov.* **10**, 47–60 (2011).
- Frömel, T. *et al.* Soluble epoxide hydrolase regulates hematopoietic progenitor cell function via generation of fatty acid diols. *Proc. Natl Acad. Sci. USA* **109**, 9995–10000 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. R. Lee, M. L. Edin and N. Gray for providing reagents; Y. Zhou, A. Dibiase, S. Yang, S. Datta, P. Manos, R. Mathieu and M. Ammerman for technical assistance; H. Huang for providing graphic illustration; R. M. White, T. E. North and C. Mosimann for discussion. Microarray studies were performed by the Molecular Genetics Core Facility at Boston Children's Hospital, supported by NIH-P50-NS40828 and NIH-P30-HD18655. S. Li in Y. Zhang's laboratory at the Longwood HMM joint core facility helped with RNA-seq. L.I.Z. and G.Q.D. are Howard Hughes Medical Institute (HHMI) investigators. This work was supported by HHMI and National Institutes of Health (NIH) grants R01 HL04880, P01P01HL32262-32, 5P30 DK49216, 5R01 DK53298, 5U01 HL10001-05, R24 DK092760, and 1R01HL097794-04 (to L.I.Z.). This work was also funded, in part, by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01 ES025034 to D.C.Z.), the National Cancer Institute grant ROCA148633-01A5 (D.P.), and DFG and Care-for-Rare Foundation (V.B.).

Author Contributions P.L. and L.I.Z. designed the study, analysed data and wrote the manuscript, with help from J.L.L. and V.B. P.L. developed the zebrafish competitive transplantation and performed the chemical screen with technical help from E.K.P. P.L. performed the mouse experiments with technical help from T.V.B., S.M. and G.C.H. P.L. performed the zebrafish microarray and embryo chemical/genetic suppressor screens with technical help from E.B.R. J.L.L. performed zebrafish embryo genetic studies and AGM timelapse imaging. V.B. performed RNA-seq and analysis on human cells with technical help from F.G.B. O.J.T. performed CHT time-lapse imaging. T.M.S. provided the chemical library. D.P. and D.C.Z. offered reagents and information related to the EET study. All authors discussed the results and commented on the manuscript.

Author Information The gene expression profiling data have been deposited in the Gene Expression Omnibus (GEO) under the accession code GSE66767. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.I.Z. (zon@enders.tch.harvard.edu).

METHODS

Zebrafish strains. Zebrafish were maintained in accordance with Animal Research Guidelines at Boston Children's Hospital (BCH). The following transgenic zebrafish were used in this study: Tg(β -actin:GFP)²³, *casper*³, RedGlo (ubiquitous *DsRed2* transgenic)²⁴, Tg(*flk1:DsRed2*)²⁵, Tg(*CD41:GFP*)²⁶, Tg(*Runx1+23:mCherry*)¹⁵ and Tg(*Runx1+23:GFP*)¹⁵. The +23 enhancer region of mouse *Runx1* was used to drive HSPC-specific expression²⁷. Tg(*flk1:dnJUNB-2A-GFP*) was constructed by cloning a human JUNBAN into a *tol2* transgenesis vector²⁸.

Chemical treatment. The ICCB Known Bioactive Library was purchased from BIOMOL (Enzo Life Sciences) and used for the adult zebrafish transplantation-based chemical screen. Chemicals were diluted at a 1:200 ratio. Chemicals used for the secondary round of screening for confirmation were from a different aliquot of the library, independent of the primary screen plate. 11,12-EET (Cayman Chemical, 50511) was resuspended in DMSO with original organic solvent evaporated. AS605240 (Sigma-Aldrich A0233) was resuspended in DMSO. The following chemicals were used for zebrafish marrow treatment: dmPGE₂ (Cayman, 14750), 10 μ M; BIO (EMD), 0.5 μ M. 0.5 μ M 11,12-EET and 14,15-EET were used for zebrafish WKM treatment (Fig. 1e); 2 μ M 11,12-EET for all mouse WBM treatment (Fig. 4); and 5 μ M 11,12-EET for all zebrafish embryo treatment (Figs 2 and 3). The concentrations were chosen based on dose titration pilot experiments with doses spanning 0.1 to 50 μ M. For the chemical suppressor screen, the suppressors were added 30 min before 11,12-EET. Zebrafish embryos were incubated with inhibitors at three different concentrations. The highest effective concentrations tested without causing general toxicity are listed in Supplementary Table 1.

Adult zebrafish kidney marrow transplantation and chemical screen. Adult zebrafish transplantation-based chemical screen was done at the human embryonic stem cell core at BCH. Three-month-old *casper* recipients (both male and female) received split-dose irradiation of 15 Gy each two days and one day before transplantation. Adult zebrafish kidney marrow cells from multiple donors were dissected, pooled together, processed into single-cell suspension and injected retro-orbitally as described previously²⁹. Tg(β -actin:GFP) WKM cells were incubated with DMSO control or chemicals in 0.9 \times DPBS plus 5% heat-inactivated FBS for 4 h at room temperature, at a density of 1,000 cells μ L⁻¹. Chemicals were washed off before 20,000 treated Tg(β -actin:GFP) WKM and 80,000 untreated RedGlo WKM were mixed together and co-injected into irradiated *casper* recipients. The number of recipients per treatment condition in the chemical screen ($n = 10$) was estimated based on preliminary experiments comparing the WKM treated with DMSO or the positive control chemical, dmPGE₂. In each experiment, recipients were randomly assigned to each treatment group. All primary hits were cherry-picked and tested in a secondary round of screening ($n = 10$ each). Recipients that died before 4 wpt, mostly owing to infection, were excluded from the analysis. No statistically significant association was observed between recipients' survival rate and a particular drug treatment.

Adult zebrafish fluorescence imaging and quantification. All zebrafish WKM transplantation results shown were obtained at 4 wpt. Transplanted adult *casper* recipients were anaesthetized with 0.2% Tricaine and imaged using a Zeiss Discovery V8 fluorescence stereomicroscope with GFP/RFP filters. To quantify the relative engraftment level in adult zebrafish, the kidney region was manually annotated for each fish, and the average fluorescence intensity of GFP and DsRed2 within the same region was measured (G_{kid} and R_{kid}) using ImageJ. The average background fluorescence intensity (G_{bkg} and R_{bkg}) was measured in a region outside the fish and a mean from multiple images within an experiment was used for all the background subtraction. The relative engraftment level was calculated as $G/R = (G_{kid} - G_{bkg}) / (R_{kid} - R_{bkg})$. The investigator analysing the data was blinded to the chemical treatment conditions. For the chemical treatment and screen results (Fig. 1e), the mean G/R in the DMSO group was normalized to 1, and all other groups were normalized to the mean G/R of DMSO. Normalized results from 2–3 independent experiments were pooled for the same chemical.

Zebrafish embryo live imaging. For live imaging, zebrafish embryos were embedded in agarose as described before^{11,15}. Single-frame images or time-lapse movies were taken on a spinning disk confocal microscope with an incubation chamber. Images of HSPC birth in the AGM were taken every 10 min. Images of the CHT engraftment process were taken every 2 min. Image post-processing and the creation of the supplementary videos were done with Fluorender, ImageJ, and Imaris.

Zebrafish embryo whole-mount *in situ* hybridization, anti-sense morpholino knockdown and mRNA overexpression. Whole-mount mRNA *in situ* hybridization experiments were performed based on the standard protocol with some modifications (http://zfinfo.org/zf_info/zfbook/chapt9/9.8.html). Embryos were scored blindly. All of the morpholinos were initially tested at 2, 4 and 6 ng to decide the effective dosage. If the morpholino did not produce a phenotype at 6 ng,

additional higher doses were tested (8, 12 ng), until the morpholino caused toxicity. See Supplementary Table 2 for morpholino sequences. *PtxA* (pertussis toxin A, Gxi inhibitor) mRNA (Addgene, plasmid 16678)³⁰ was *in vitro* transcribed with SP6 RNA polymerase (Ambion, mMESSAGE mMACHINE SP6, AM1340) and injected into one-cell stage zebrafish embryos at 3 pg per embryo, causing morphological defects but no general toxicity.

Zebrafish embryo proliferation and apoptosis assays. Zebrafish embryos were chemically treated between 48 and 72 hpf, and fixed at 72 hpf. For proliferation analysis, embryos were permeabilized and stained with primary antibody against phospho-histone H3, and FITC-conjugated secondary antibody. Embryos were imaged and phospho-H3-positive cells in the CHT were manually counted. Secondary antibody-only control showed no nonspecific staining. For apoptosis analysis, embryos were stained using the colorimetric TUNEL staining kit (Promega).

Cell culture. Human CD34⁺ cells were isolated from fresh umbilical cord blood by Ficoll separation of mononuclear cells and subsequent positive selection of CD34⁺ cells using magnetic beads (Miltenyi). Cells were treated in serum-free IMDM media (Sigma-Aldrich) with either DMSO or 5 μ M 11,12-EET for 2 h at 37 °C. U937 cells³¹ were cultured in RPMI-1640 Medium (Sigma-Aldrich) and 10% FBS at 5% CO₂ in air atmosphere according to the protocol (purchased from ATCC without additional confirmation or examination for mycoplasma contamination). For *in vitro* treatment, cells were serum-starved for 1 h and then treated with either DMSO or 5 μ M 11,12-EET for 2 h at 37 °C. The conditions for use of human umbilical cord blood CD34⁺ cells are governed by the associated institution's Internal Review Board (IRB) on behalf of the DF/HCC in accordance with Department of Health and Human Services regulations at 45 CFR Part 46. Informed consent was obtained from all subjects.

Mouse bone marrow transplant. All mice were maintained according to IACUC approved protocols in accordance with BCH animal research guidelines. Nine-week-old CD45.1 and CD45.2 (C57/BL6) male mice were purchased from Jackson Laboratories and housed for 2–3 weeks before the experiments. All CD45.2 recipients received an 11 Gy split dose of γ -irradiation before transplantation, and were randomly assigned to each treatment group. 20,000 CD45.1 WBM cells from age- and gender-matched BL6 donors were treated in DMEM plus 2% FBS at room temperature for 4 h with 2 μ M 11,12-EET. For the suppressor experiment (Fig. 4f), 10 μ M LY294002 was added to the cells 30 min before the addition of 11,12-EET. Chemicals were washed off before cells were resuspended in PBS and mixed with 200,000 fresh CD45.2 mouse WBM cells. Donor cells were retro-orbitally injected into CD45.2 recipients. Each treatment condition included 10 recipients per experiment. The 12-week survival rate in each experiment was 90–95%, and recipients that died before 12 wpt were excluded from the analysis.

Mouse peripheral blood chimaerism analysis. Peripheral blood was stained with lineage-specific antibodies and analysed on LSRII (BD Biosciences) to assess engraftment. The following antibodies were used: Gr1 (RB6-8C5), Mac1 (M1/70), B220 (RA3-B2), CD3 (145-2C11) and Ter119 from eBioscience; CD45.1 and CD45.2 from BD Biosciences. The CD45.1 chimaerisms in non-irradiated, untransplanted CD45.2 mice were used as a negative staining control. Recipients with multi-lineage chimaerism above the average negative-control chimaerism plus 3 standard deviations were considered to have multi-lineage engraftment (Fig. 4f).

Mouse competitive homing assay. The mouse competitive homing experiment was performed as described, with modifications³². In brief, CD45.1 mouse WBM were treated with either DMSO or 2 μ M 11,12-EET at room temperature for 3.5 h at a density of 2×10^6 cells per ml. DiO dye was added to the cell suspension (1:200) and incubated at 37 °C for 30 min. At the same time, WBM from CD45.2 mice were incubated at room temperature for 3.5 h without chemical treatment, then labelled with DiD dye (1:200) at 37 °C for 30 min. After the incubation and labelling, the chemicals and dyes were washed off. The DiO-labelled CD45.1 bone marrow and DiD-labelled CD45.2 WBM were mixed at a 1:1 ratio and competitively transplanted into CD45.2 recipients (2.5×10^6 from each donor). Recipients received total body irradiation of 11 Gy one day before transplantation. 16 h after transplant, the recipients were euthanized and bone marrow was analysed by flow cytometry for both DiO/DiD and surface lineage markers (Gr1, Mac1, B220, CD3, Ter119, from eBioscience) and c-Kit (2B8, BD Biosciences). The ratio between the percentages of DiO⁺ (donor) and DiD⁺ (competitor) cells within different cell populations was quantified. DiO and DiD are from Vybrant Multicolor Cell-Labeling Kit (Molecular Probes, V-22889).

Mouse bone marrow apoptosis and proliferation assays. For apoptosis analysis, mouse WBM cells were treated with DMSO or 2 μ M 11,12-EET for 4 h *in vitro* and stained using the AnnexinV apoptosis kit (BD Biosciences), together with antibodies against lineage markers, Sca-1 (E13-161.7) and c-Kit (2B8). The 7-AAD⁺/annexinV⁺ cells are the apoptotic population. For proliferation analysis, mouse WBM were treated with DMSO or 2 μ M 11,12-EET for 4 h *in vitro*, in the presence

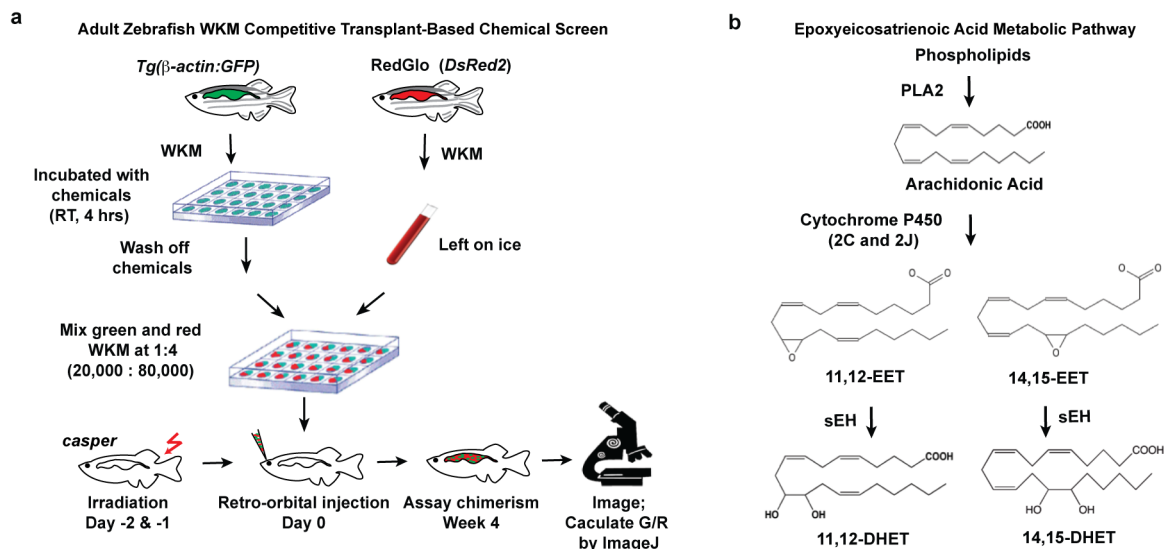
of 10 μ M BrdU, then fixed, permeabilized and stained with anti-BrdU antibody (BD Pharmingen BrdU Flow Kits)³³, together with antibodies against lineage markers, Sca-1 and c-Kit.

Gene expression profiling and IPA analysis. Gene expression profiling data are available in GEO (accession numbers GSE39707 and GSE66767). For the zebrafish embryo gene expression study, total RNA was extracted from 36 hpf zebrafish embryos treated with DMSO or 5 μ M 11,12-EET between 24 and 36 hpf, with three biological replicates each and $n = 25$ in each group. Microarray hybridization was performed with the Affymetrix GeneChip Zebrafish Genome Array. Hybridized microarray was background-corrected, normalized and multiple-tested using Goldenspike (<http://www2.ccr.buffalo.edu/halfon/spike/>) in R/Bioconductor³⁴. Genes with $q < 0.1$ by SNR test were considered differentially expressed (Supplementary Table 3). For RNaseq analysis on human cells, total RNA was extracted from treated CD34⁺ and U937 cells with the RNeasy mini plus kit from Qiagen. After quality control on the Bioanalyzer (Agilent), total RNA was depleted of ribosomal RNA with the RiboZero gold kit (Epicentre). Enriched mRNA was applied to library preparation according to manufacturer's protocol (NEBNext Ultra). After repeated quality control for average DNA input size of 300 base pairs (bp), samples were sequenced on a HiSeq Illumina sequencer with 2×100 -bp paired-end reads. Quality control of RNA-Seq data sets was performed by FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and Cutadapt³⁵ to remove adaptor sequences and low quality regions. The high-quality reads were aligned to UCSC build hg19 of the human genome using Tophat 2.0.11 without novel splicing form calls³⁶. Transcript abundance and differential expression were calculated with Cufflinks 2.2.1 (ref. 37). FPKM values were used to normalize and quantify each transcript. $\log_2(fc)$ (\log_2 fold change), P and q values were calculated. As the experiment was not performed in biological replicates, the P and q values were not taken into consideration for further analysis of the data. Results are listed with a cutoff of $\log_2(fc) > 0.5$ for upregulated genes and $\log_2(fc) < -0.5$ for downregulated genes in Supplementary Table 4. Analysis of overlapping upregulated genes in both cell types after EET treatment was done using Venny (<http://bioinfogp.cnb.csic.es/tools/venny/index.html>). The list of overlapping genes was analysed using IPA (QIAGEN) to map enriched bio-functions.

Statistics. The comparison of multi-lineage engraftment in Fig. 4b and f were done by two-tailed Fisher's exact test by comparing the number of engrafted versus non-engrafted recipients. Using the mean chimaerism plus $2 \times$ s.e.m. in the DMSO control group as the cutoff, recipients with a chimaerism higher than the cutoff were considered engrafted (Fig. 4b). Embryos in the *in situ* hybridization experiments were scored blindly and analysed by Chi-square tests or two-tailed Fisher's exact test in the case of small sample sizes. The rest of the statistics were done with

unpaired two-tailed t -test. Graphs show mean with s.e.m. No statistical methods were used to predetermine sample size. All the zebrafish embryos, adult zebrafish and mice for transplantation were randomized into each treatment group.

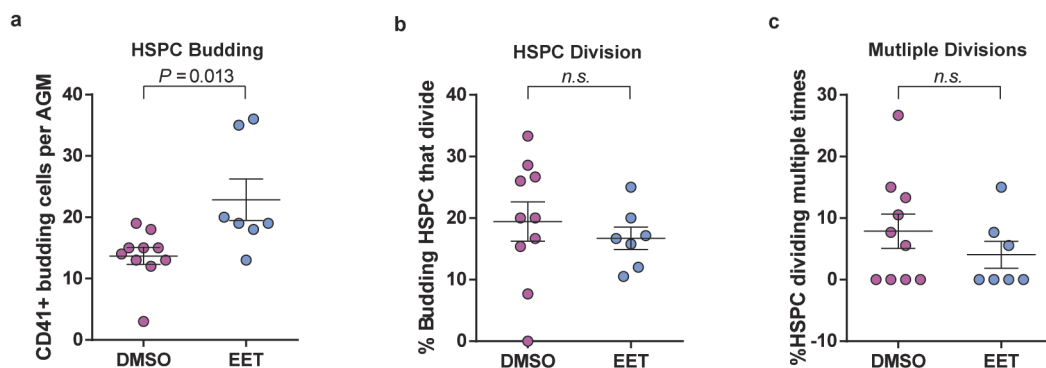
23. Traver, D. *et al.* Transplantation and *in vivo* imaging of multilineage engraftment in zebrafish bloodless mutants. *Nature Immunol.* **4**, 1238–1246 (2003).
24. Blake, A., Crockett, R., Essner, J., Hackett, P. & Nasevicius, A. Recombinant constructs and transgenic fluorescent ornamental fish therefrom. US patent US7,700,825 B2 (2010).
25. Kikuchi, K. *et al.* Retinoic acid production by endocardium and epicardium is an injury response essential for zebrafish heart regeneration. *Dev. Cell* **20**, 397–404 (2011).
26. Ma, D., Zhang, J., Lin, H. F., Italiano, J. & Handin, R. I. The identification and characterization of zebrafish hematopoietic stem cells. *Blood* **118**, 289–297 (2011).
27. Bee, T. *et al.* The mouse *Runx1* +23 hematopoietic stem cell enhancer confers hematopoietic specificity to both *Runx1* promoters. *Blood* **113**, 5121–5124 (2009).
28. Ikebe, D., Wang, B., Suzuki, H. & Kato, M. Suppression of keratinocyte stratification by a dominant negative JunB mutant without blocking cell proliferation. *Genes Cells* **12**, 197–207 (2007).
29. Pugach, E. K., Li, P., White, R. & Zon, L. Retro-orbital injection in adult zebrafish. *J. Vis. Exp.* **34**, 1645 (2009).
30. Slusarski, D. C., Corces, V. G. & Moon, R. T. Interaction of Wnt and a Frizzled homologue triggers G-protein-linked phosphatidylinositol signalling. *Nature* **390**, 410–413 (1997).
31. Sundstrom, C. & Nilsson, K. Establishment and characterization of a human histiocytic lymphoma cell line (U-937). *Int. J. Cancer* **17**, 565–577 (1976).
32. Lam, B. S., Cunningham, C. & Adams, G. B. Pharmacologic modulation of the calcium-sensing receptor enhances hematopoietic stem cell lodgment in the adult bone marrow. *Blood* **117**, 1167–1175 (2011).
33. Challen, G. A., Boles, N., Lin, K. K. & Goodell, M. A. Mouse hematopoietic stem cell identification and analysis. *Cytometry A* **75**, 14–24 (2009).
34. Choe, S. E., Boutros, M., Michelson, A. M., Church, G. M. & Halfon, M. S. Preferred analysis methods for Affymetrix GeneChips revealed by a wholly defined control dataset. *Genome Biol.* **6**, R16 (2005).
35. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 1 (2011).
36. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
37. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.* **28**, 511–515 (2010).
38. Lee, C. R. *et al.* Endothelial expression of human cytochrome P450 epoxygenases lowers blood pressure and attenuates hypertension-induced renal injury in mice. *FASEB J.* **24**, 3770–3781 (2010).



Extended Data Figure 1 | Zebrafish WKM competitive transplantation-based chemical screen identifies EETs as enhancers of marrow engraftment.

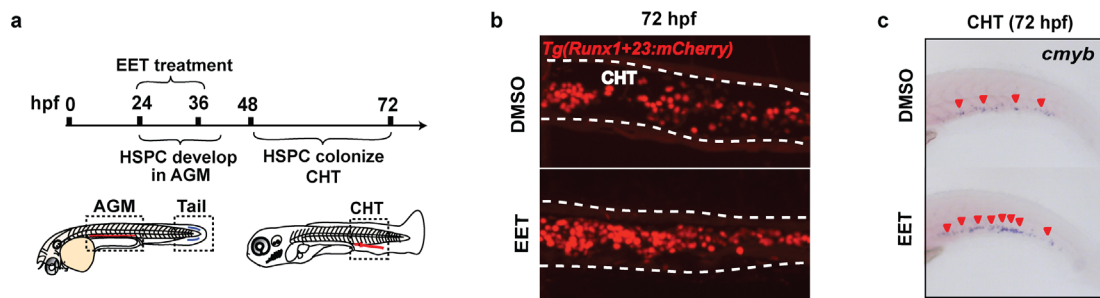
a, WKM from *Tg(β-actin:GFP)* donors were dissected, dissociated as single-cell suspension, and incubated with chemicals at room temperature for 4 h in a round-bottom 96-well plate. Meanwhile, WKM were dissected from RedGlo zebrafish, counted and kept on ice. After the drug treatment, chemicals were washed off and cells were resuspended in 0.9× PBS plus 5% FBS. Approximately 20,000 treated green WKM and 80,000 untreated red WKM were co-injected retro-orbitally into sublethally irradiated *casper* zebrafish ($n = 10$ per chemical). For every independent screening day, negative control

(DMSO) and positive control (10 μM dmPGE₂) treatments were used for normalization and quality assurance. The engraftment was measured at 4 wpt by fluorescence imaging and ImageJ quantification as described in Fig. 1b. **b**, EET metabolic pathway: arachidonic acid is released by phospholipase A₂ (PLA₂) from the membrane lipid bilayer. EETs are synthesized directly from arachidonic acid by the cytochrome P450 family of epoxygenases, especially 2C and 2J in human³⁸, and get degraded by soluble epoxide hydrolase (sEH), generating dihydroxyeicosatrienoic acids (DiHET). Four isomers of EET exist *in vivo*: 5,6-, 8,9-, 11,12- and 14,15-EET.



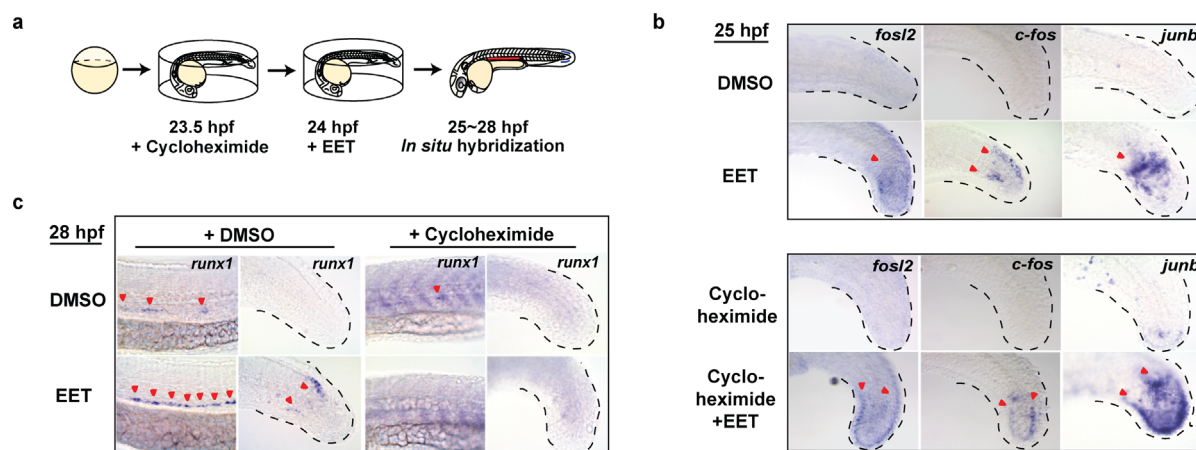
Extended Data Figure 2 | 11,12-EET enhances HSPC specification in the AGM in zebrafish embryos. Tg(*CD41:GFP/flk1:DsRed2*) embryos were treated with DMSO or 5 μ M 11,12-EET starting at 24 hpf, then mounted for spinning disc confocal timelapse imaging from 30–46 hpf in the presence of the chemicals. Data are mean and s.e.m., unpaired two-tailed *t*-tests, $n = 10$ for DMSO, $n = 7$ for EET. **a**, More HSPCs are directly specified in

EET-treated AGM. Graph shows HSPCs born by direct specification/budding only, excluding cells born by division of an already-budding cell. **b**, **c**, 11,12-EET does not influence the rate of HSPC division in the AGM, shown by per movie, percentage of budding HSPCs that divide at least once (**b**) and divide twice or more (**c**) before leaving the AGM or before the end of timelapse recording.



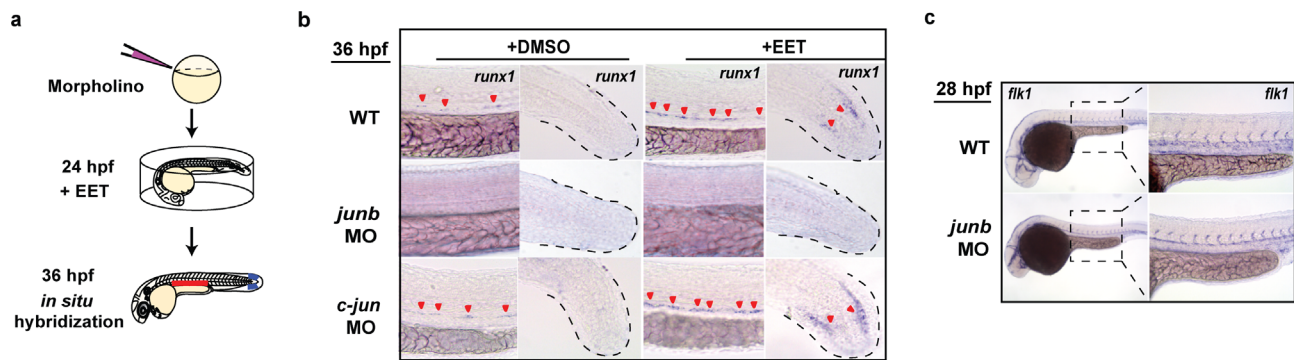
Extended Data Figure 3 | 11,12-EET treatment between 24 and 48 hpf increases the number of HSPCs in the CHT. **a**, Embryos were treated between 24 and 48 hpf with either DMSO or 5 μ M 11,12-EET. Chemicals were washed off at 48 hpf, and embryos grew in drug-free environment for another 24 h. **b**, 11,12-EET treatment increased the number of mCherry⁺ HSPCs in the CHT

in *Tg(Runx1+23:mCherry)* embryos (see also Fig. 2e). Representative images of the CHT from the two groups. **c**, The same chemical treatment increased the staining of *cmyb*, a HSPC marker, by whole-mount RNA *in situ* hybridization. Representative images from each group (a total of $n > 60$ from three independent experiments).



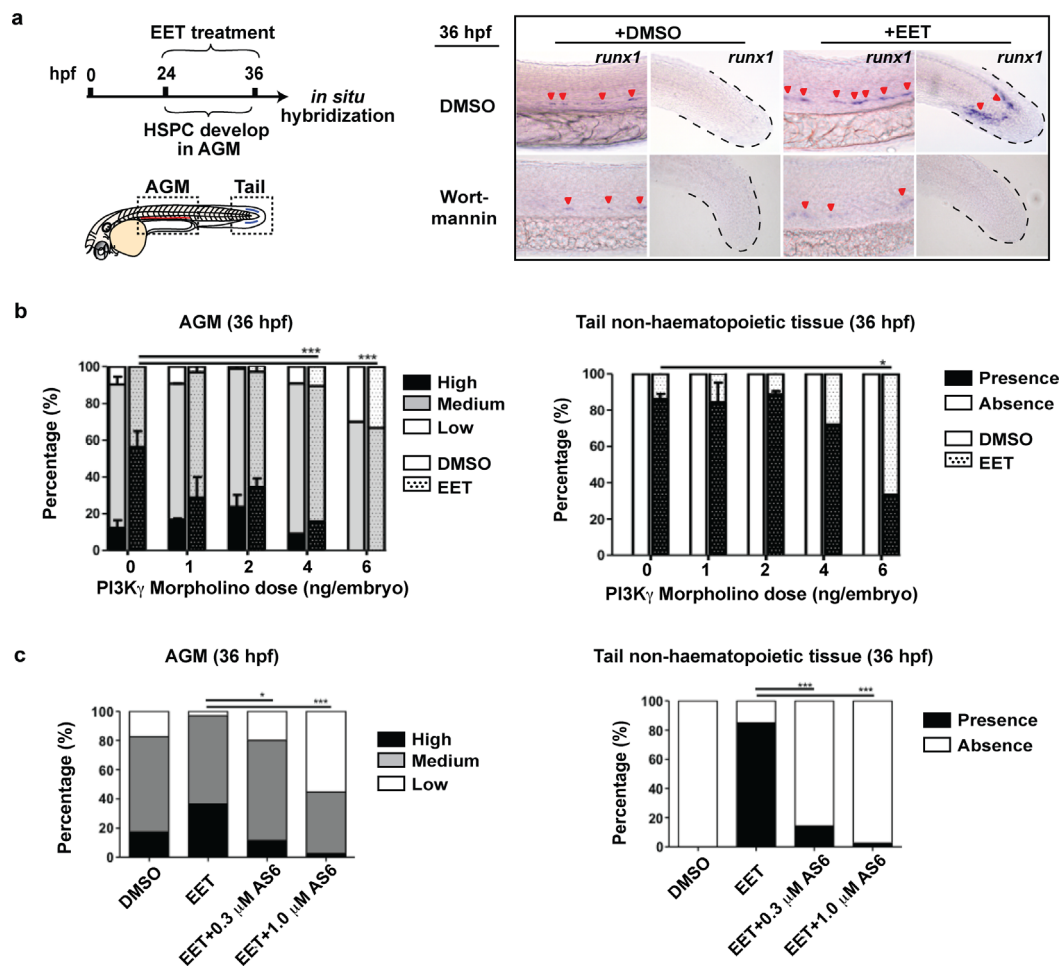
Extended Data Figure 4 | EET signalling pathway activates AP-1 family members as primary transcriptional targets, and *runx1* as a secondary transcriptional target. **a**, Wild-type embryos were incubated with 300 μ M cycloheximide, a translation blocker, for 30 min before the addition of 5 μ M 11,12-EET at 24 hpf. Embryos were fixed for *in situ* hybridization at 25 hpf or 28 hpf. **b**, AP-1 transcription was induced after 1 h treatment with 11,12-EET, insensitive to cycloheximide inhibition. This means AP-1 induction does not depend on *de novo* protein synthesis, indicating AP-1 members are primary

transcriptional targets of the EET signalling pathway. **c**, *runx1* transcription was induced after 4 h treatment with EET (two columns on the left) and cycloheximide completely blocked EET-induced *runx1* expression (two columns on the right). This suggests *runx1* transcription depends on *de novo* protein synthesis of an upstream factor(s) upon EET stimulation, indicating that *runx1* is a secondary transcriptional target of the EET signalling pathway. Representative images from each group (a total of $n > 30$ from two independent experiments).



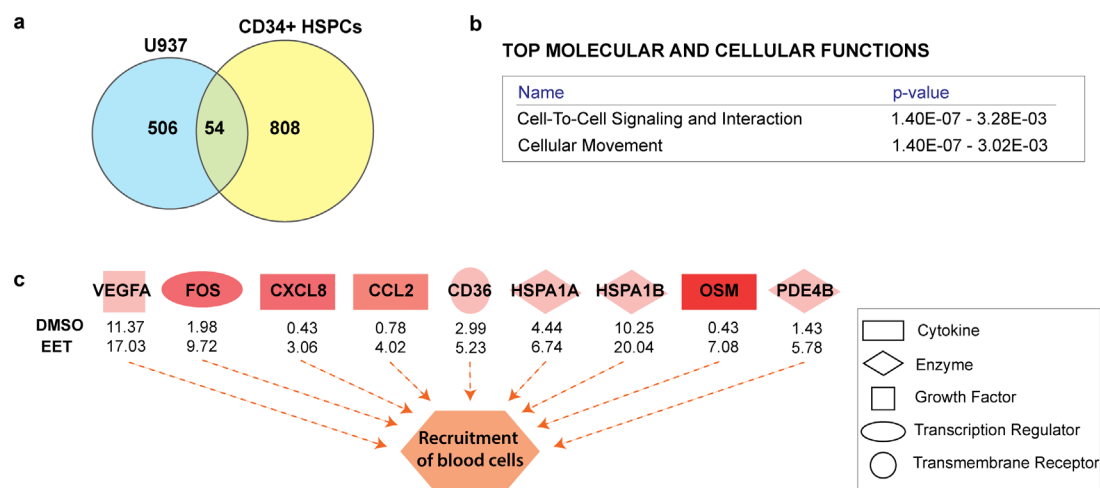
Extended Data Figure 5 | Knocking down *junb* and *junbl* inhibits HSPC specification in the AGM. **a**, Wild-type embryos were injected with antisense morpholinos at the one-cell stage, and treated with DMSO or 5 μ M 11,12-EET starting from 24 hpf. Embryos were fixed at 36 hpf for *in situ* hybridization of *runx1*. **b**, Knocking down *junb* completely blocked *runx1* expression at 36 hpf both in the AGM and the tail non-haematopoietic tissue (middle row).

By contrast, knocking down *c-jun* did not block the increase of *runx1* (bottom row), consistent with the lack of *c-jun* upregulation in EET-treated embryos (data not shown). **c**, *junb* morphants still developed normal vascular structure in the AGM at 28 hpf, as shown by endothelial marker *flk1*. Representative images from each group (a total of $n > 40$ from three independent experiments).



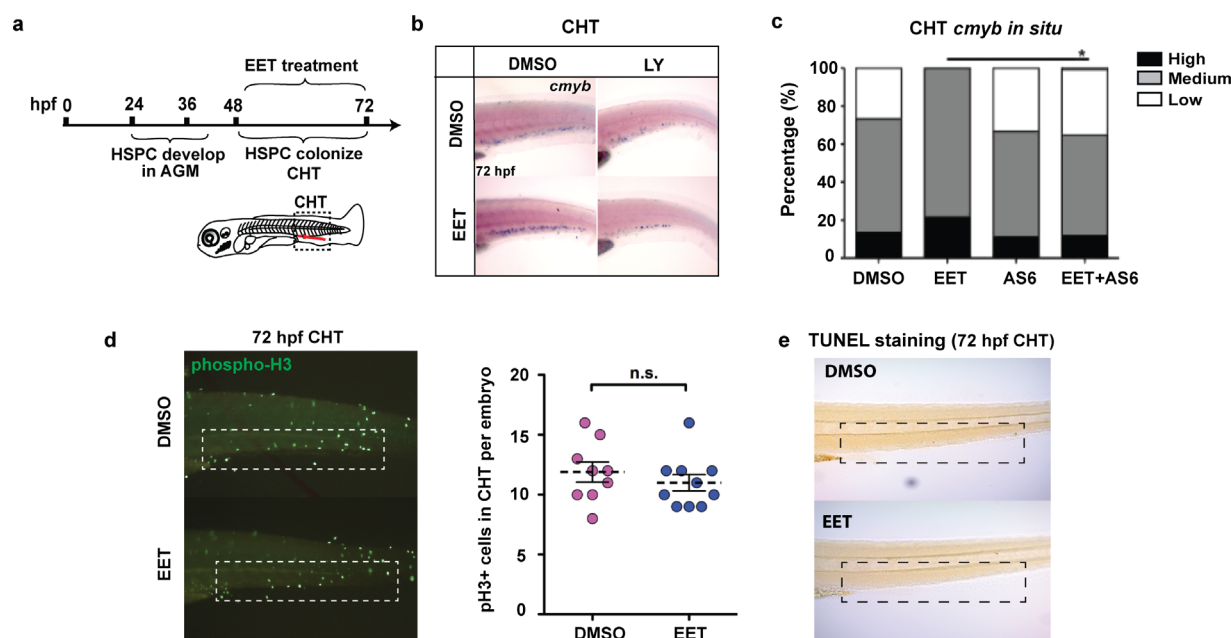
Extended Data Figure 6 | PI(3)K γ activation is specifically required for EET-induced gene expression signature. **a**, Similar to LY294002 (Fig. 3d–e), another pan-PI(3)K/AKT inhibitor, wortmannin (1 μ M), blocked EET-induced *runx1* expression both in the AGM and tail. Representative images from each group (a total of $n > 60$ from three independent experiments). **b**, Morpholinos specific to PI(3)K γ , but not α , β and δ subunits (data not shown), prevented EET-induced *runx1* in the AGM and tail. Embryos were injected at 1–2-cell stage with the indicated amount of morpholino and treated with DMSO or 5 μ M 11,12-EET from 24–36 hpf. *In situ* hybridization for *runx1* performed at 36 hpf and percentages of embryos having high, medium or

low expression in the AGM and present or absent expression in the tail are shown. Graph summarizes three experiments, $n \geq 10$ embryos for each condition (0, 1 and 2 ng, data are mean and s.e.m.) or one experiment $n \geq 9$ for all conditions (4 and 6 ng). **c**, The PI(3)K γ -specific inhibitor AS605240 (AS6) recapitulates the morpholino phenotype. Embryos treated from 24 to 36 hpf with DMSO or 5 μ M 11,12-EET, with or without 0.3–1.0 μ M AS6, then fixed and stained for *runx1* at 36 hpf. DMSO, $n = 23$; EET, $n = 33$; EET+0.3 μ M AS6, $n = 35$; EET+1.0 μ M AS6, $n = 38$. * $P < 0.05$, *** $P < 0.001$, two-tailed Fisher's exact test.



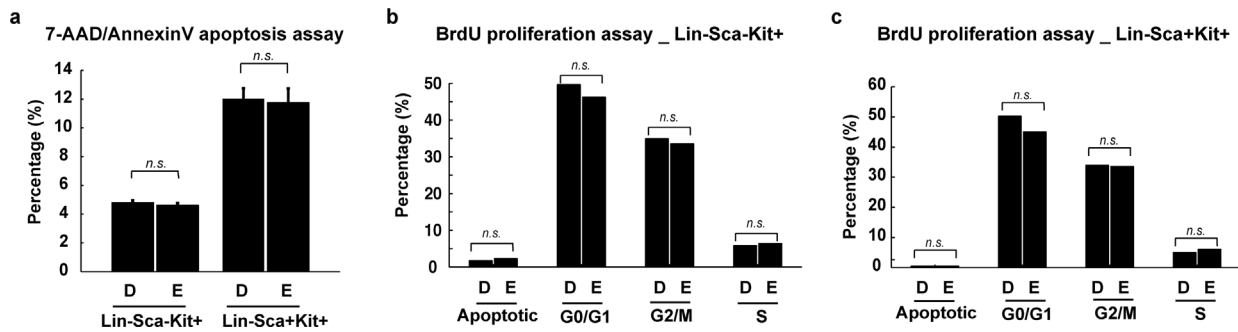
Extended Data Figure 7 | 11,12-EET upregulates genes involved in cell-to-cell signalling and cellular movement in haematopoietic progenitors.
a, Venn diagram showing a common set of 54 genes upregulated ($\log_2(\text{fc}) > 0.5$) after 2 h of 11,12-EET treatment (5 μM), both in human myeloid U937 cells and human umbilical cord CD34⁺ HSPCs (see also Supplementary Table 4 for lists of up- and downregulated genes). **b**, **c**, Ingenuity Pathway Analysis (IPA) of the overlapping gene set between the two cell types for enrichment of bio-functions. **b**, Biological processes, such as cell-to-cell signalling and cellular movement, were highly enriched, supporting the capability of EETs in

enhancing engraftment (see also Supplementary Table 4 for a comprehensive list of all biological functions predicted to be activated or suppressed based on the same gene set). **c**, Activation of recruitment of blood cells is caused by upregulation of chemokines and cytokines such as CXCL8 and OSM after EET treatment, as well as by upregulation of transcription factors, such as AP-1 genes (FOS). Orange dashed arrows depict activation. Shades of red represent the level of activation. Numbers underneath factors show RNaseq FPKM (fragments per kilobase of exon per million reads mapped) values in U937 cells.



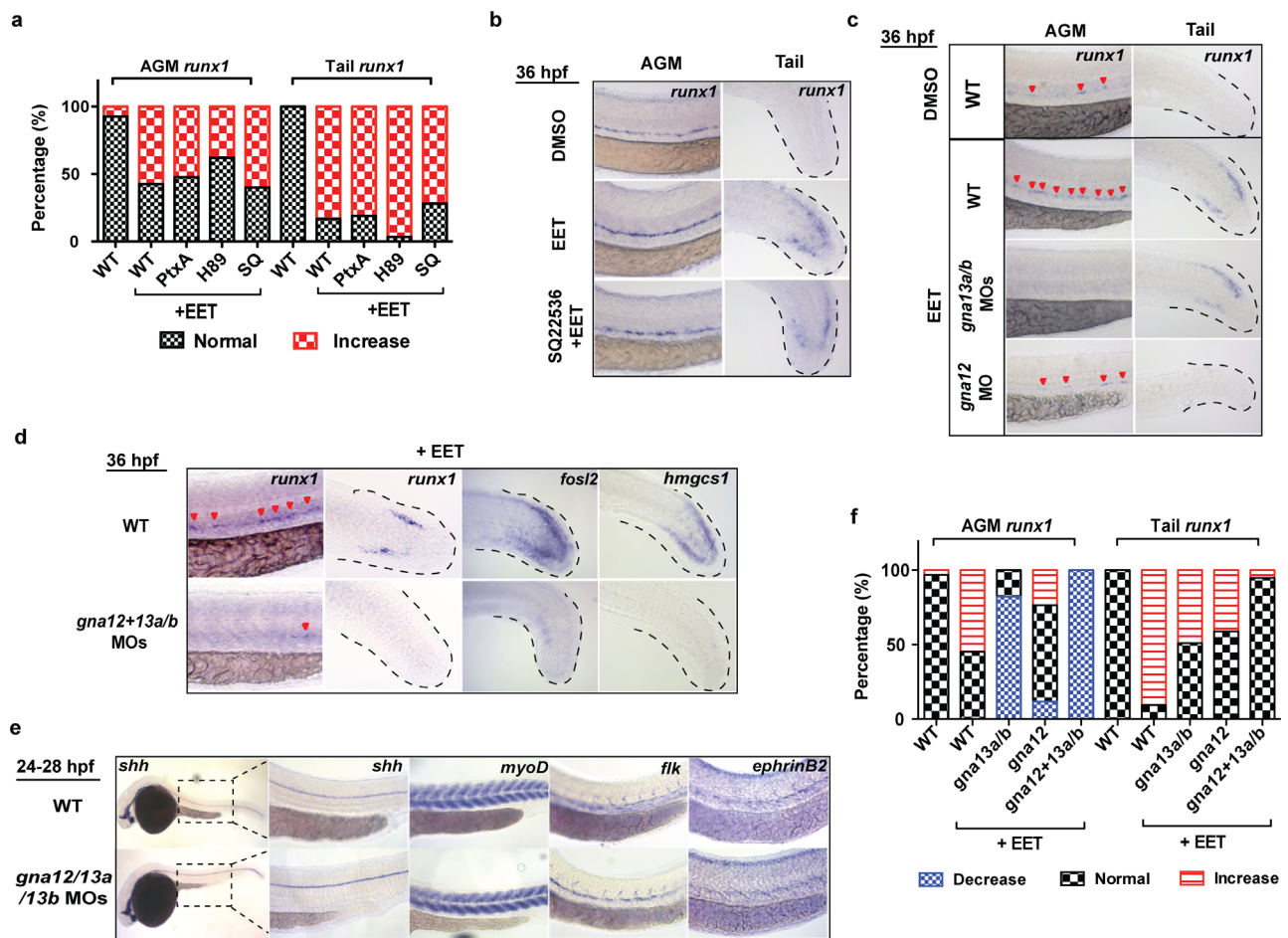
Extended Data Figure 8 | 11,12-EET treatment after HSPC specification still enhances the number of HSPCs in the CHT. **a**, Embryos were treated with DMSO or 5 μ M 11,12-EET between 48 and 72 hpf to bypass the HSPC specification process in the AGM. 72-hpf embryos were fixed and tested on the following assays. **b**, *In situ* hybridization for *cmyb*, a marker for HSPCs. EET treatment significantly increased the staining, while LY294002, a pan-PI(3)K inhibitor, suppressed the effect. Representative images from each group (a total of $n > 60$ from four independent experiments). **c**, A PI(3)K γ -specific inhibitor AS605240 (AS6) also blocked the EET-induced increase of *cmyb* staining. Percentage of embryos having high, medium or low expression in the CHT is

shown. $n \geq 11$ for all conditions. Chi-square analysis. **d**, The increase of HSPCs in the CHT is not due to effects on proliferation. Immunofluorescence staining for phospho-histone H3 (pH3) as a marker for proliferating cells. The number of pH3-positive cells was manually counted. Two-tailed *t*-test showed no significant difference between DMSO- versus EET-treated embryos. $n = 9$ for DMSO, $n = 10$ for EET. **e**, TUNEL staining as an assay for apoptotic cells. Apoptosis was minimal in the CHT at 72 hpf. As a staining control, obvious apoptosis was detected in the same embryos in the brain region, and was comparable between DMSO- and EET-treated embryos (data not shown).



Extended Data Figure 9 | 11,12-EET treatment of mouse WBM does not lead to immediate changes in cell proliferation or apoptosis. **a**, *In vitro* apoptosis assay on WBM treated with DMSO or 2 μ M 11,12-EET for 4 h. The 7-AAD-negative and annexinV-positive population are the cells undergoing apoptosis. No significant differences between the two groups were observed either in Lin⁻Sca⁻Kit⁺ or Lin⁻Sca⁺Kit⁺ progenitor populations ($n = 4$ each),

mean and s.e.m. **b**, **c**, *In vitro* proliferation assay on WBM treated with DMSO or 2 μ M 11,12-EET for 4 h, in the presence of 10 μ M BrdU. No significant differences between the two groups were observed either in Lin⁻Sca⁻Kit⁺ (**b**) or Lin⁻Sca⁺Kit⁺ populations (**c**) for any cell cycle stage. Unpaired *t*-test, $n = 4$ each, bar denotes the mean. D, DMSO; E, EET.



Extended Data Figure 10 | *Ga12/13* is specifically required for EET-induced phenotypes in zebrafish embryos. All embryos were treated with DMSO or 5 μ M 11,12-EET between 24 and 36 hpf. Chemical inhibitors were added 30 min before EET. mRNA or morpholinos (MO) were injected at the one-cell stage. **a, b**, Inhibiting $G\alpha s$ or $G\alpha i$ had no effect on EET-induced *runx1* expression. Embryos were categorized into two groups with either normal or increased *runx1* expression level ($n > 20$ each). PtxA, pertussis toxin A, 3 pg, inhibiting $G\alpha i$ (ref. 30); H89, 5 μ M, PKA inhibitor downstream of $G\alpha s$; SQ, SQ22536, 50 μ M, adenylate cyclase inhibitor downstream of $G\alpha s$. Representative images from each group (**b**) (a total of $n > 40$ from two independent experiments). **c–f**, Synergistic effects of *gna12/13a/13b*

knockdown on suppressing *runx1* expression. Knocking down *gna13a/b* or *gna12* alone partially inhibited EET-induced *runx1* expression in the AGM and tail (**c**). *gna12* MO: 2 ng; *gna13a/13b* MOs: 1 ng each. Triple morpholinos against *gna12*, *gna13a* and *gna13b* (0.67 ng each) completely blocked EET-induced multiple gene expression, including *runx1*, genes in regeneration (*fosl2*) and cholesterol metabolism (*hmgcs1*) (**d**), while other major tissue development processes were not significantly affected, such as notochord (*shh*), muscle (*myoD*), and blood vessels (*flk*, *ephrinB2*) (**e**). **f**, The results were quantified. Embryos were categorized as having decreased, normal or increased *runx1* expression. The bar graph represents the percentage of embryos in each group ($n > 30$).

Redox rhythm reinforces the circadian clock to gate immune response

Mian Zhou^{1,2*†}, Wei Wang^{1,2*†}, Sargis Karapetyan³, Musoki Mwimba^{1,2}, Jorge Marqués^{1,2}, Nicolas E. Buchler^{2,3} & Xinnian Dong^{1,2}

Recent studies have shown that in addition to the transcriptional circadian clock, many organisms, including *Arabidopsis*, have a circadian redox rhythm driven by the organism's metabolic activities^{1–3}. It has been hypothesized that the redox rhythm is linked to the circadian clock, but the mechanism and the biological significance of this link have only begun to be investigated^{4–7}. Here we report that the master immune regulator NPR1 (non-expressor of pathogenesis-related gene 1) of *Arabidopsis* is a sensor of the plant's redox state and regulates transcription of core circadian clock genes even in the absence of pathogen challenge. Surprisingly, acute perturbation in the redox status triggered by the immune signal salicylic acid does not compromise the circadian clock but rather leads to its reinforcement. Mathematical modelling and subsequent experiments show that NPR1 reinforces the circadian clock without changing the period by regulating both the morning and the evening clock genes. This balanced network architecture helps plants gate their immune responses towards the morning and minimize costs on growth at night. Our study demonstrates how a sensitive redox rhythm interacts with a robust circadian clock to ensure proper responsiveness to environmental stimuli without compromising fitness of the organism.

Life on Earth has evolved the circadian clock to anticipate diurnal and seasonal changes⁸. This 'scheduling' mechanism coordinates biological processes to reduce random energy expenditures and increase fitness. In *Arabidopsis*, daily time-keeping is driven by three interlocked transcription–translation feedback loops (TTFLs): the core loop, the morning loop, and the evening loop. The core loop consists of three transcription factors: two partly redundant morning-phased CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) and LATE ELONGATED HYPOCOTYL (LHY), and the evening-phased TIMING OF CAB2 EXPRESSION 1 (TOC1). CCA1/LHY and TOC1 are repressors of each other's expression^{9,10}. Besides the TTFL circadian clock, non-transcriptional redox oscillations exist in all domains of life, including *Arabidopsis*³. Even though redox rhythm was shown to influence the TTFL clock³, how these two oscillatory systems are linked molecularly, and what the biological significance of having two oscillatory systems is, remain largely unknown.

To begin addressing these questions, we examined the daily changes in the reduction–oxidation coenzymes NADPH and NADP⁺ in *Arabidopsis* under constant light and found them to display circadian rhythms ($P < 10^{-4}$), with NADPH peaking before subjective dawn and NADP⁺ peaking before subjective dusk (Fig. 1a, b). Moreover, their ratio also oscillated in a circadian manner (Extended Data Fig. 1). These data support the existence of widespread metabolic and redox rhythms in plants beyond the previously reported oscillations of oxidized peroxiredoxin, H₂O₂, and catalases^{3,7,11}. It is known that the plant immune-inducing signal salicylic acid (SA) can alter the cellular redox to trigger defence gene expression¹². We found that under constant

light, treating plants with SA could significantly perturb NADPH and NADP⁺ rhythms as well as their ratio (Fig. 1a, b and Extended Data Fig. 1), indicating that the redox rhythm is sensitive to external perturbations.

We next examined whether this SA-triggered redox rhythm perturbation could be transduced to the circadian clock by first focusing on the evening-phased *TOC1*, which is responsive to many environmental factors¹³. Using quantitative PCR (qPCR), we observed significant increases in amplitude and average expression of *TOC1* upon SA treatment (Fig. 1c). Similar results were observed using a transgenic line carrying a reporter of the *TOC1* promoter fused to luciferase (*TOC1p:LUC*)¹⁴ (Fig. 1d, e and Extended Data Fig. 2a). Strikingly, the period of the *TOC1p:LUC* expression rhythm did not change, regardless of whether SA was applied at subjective dawn (Fig. 1d) or dusk (Fig. 1e).

To study the effect of endogenous SA, which oscillates in a circadian manner¹⁵, on the clock, we crossed the *TOC1p:LUC* reporter into the SA biosynthesis mutant, *sid2* (SA induction-deficient 2)¹⁶. We found that the amplitude and the average expression of *TOC1* were significantly reduced in *sid2* and this phenotype was rescued upon treatment with exogenous SA (Extended Data Fig. 2b). Our results indicate that endogenous SA plays a part in the redox rhythm that modulates the amplitude and average expression of the circadian clock.

SA-induced redox changes can lead to reduction of the master immune regulator, NPR1, the release of NPR1 monomer for nuclear translocation, defence gene induction¹², and subsequent degradation mediated by the nuclear SA receptors NPR3 and NPR4 (ref. 17). To test whether the SA-mediated regulation of *TOC1* is through NPR1, we crossed *TOC1p:LUC* into the *npr1* mutant¹⁸. We found that the mutation not only dampened the basal expression of *TOC1* but also abolished the SA-triggered increases in expression regardless of the time of treatment (Fig. 2a and Extended Data Fig. 3a–c).

We hypothesized that NPR1 is an intrinsic regulator of *TOC1* in response to the rhythmic accumulation of the endogenous SA¹⁵. Through western blotting, we indeed found a circadian oscillatory pattern for the NPR1 monomer ($P < 0.01$) with a peak at night (Fig. 2b and Extended Data Fig. 4a). Therefore, oscillation in the endogenous SA level may drive the rhythmic nuclear translocation of NPR1 to regulate the circadian clock genes. To test this hypothesis, we used mutants of cytoplasmic-localized thioredoxins (TRX), *trx-h3* and *trx-h5*, in which NPR1 nuclear translocation is largely impaired¹⁹. We found that both the basal rhythm of *TOC1p:LUC* and its responsiveness to SA were diminished in *trx-h3 trx-h5* (Fig. 2c and Extended Data Fig. 5a), suggesting the requirement of NPR1 nuclear translocation in regulating *TOC1* expression. Besides SA, glutathione-reduced ethyl ester (GSH_{me}), a redox-altering reagent²⁰, could also enhance *TOC1* expression in an NPR1-dependent manner (Extended Data Fig. 5b), suggesting that NPR1 is a general redox sensor in modulating this clock gene.

¹Howard Hughes Medical Institute–Gordon and Betty Moore Foundation, Duke University, Durham, North Carolina 27708, USA. ²Department of Biology, PO Box 90338, Duke University, Durham, North Carolina 27708, USA. ³Department of Physics, Duke University, Durham, North Carolina 27708, USA. [†]Present address: Department of Plant Pathology and Microbiology, Iowa State University, 415 Bessey Hall, Ames, Iowa 50011, USA.

*These authors contributed equally to this work.

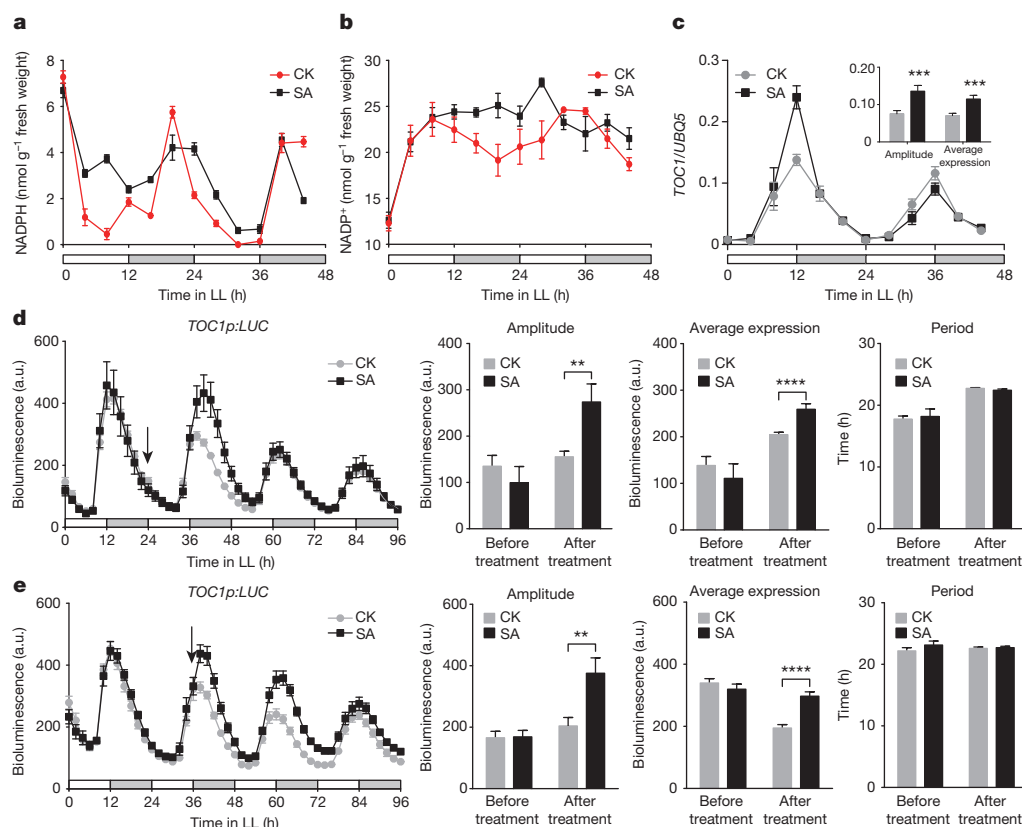


Figure 1 | SA disrupts redox rhythm but boosts *TOC1* expression without changing its period. **a–c**, NADPH (**a**), NADP⁺ (**b**), and *TOC1* messenger RNA (mRNA) (**c**) in plants after application of water (CK) or SA at 0 h under constant light (LL). White and grey bars represent subjective days and nights, respectively. Data are mean \pm s.e.m. ($n = 3$; t -test; *** $P < 0.001$).

NPR1 is a transcription cofactor of the TGA class of transcription factors in SA-induced defence gene expression²¹. Using a yeast one-hybrid assay, six *Arabidopsis* TGAs were found to have strong binding affinities to the *TOC1* promoter at the two TGA-binding sites (TBS) (Fig. 2d). To confirm this *in planta*, we mutated TBS in the *TOC1p:LUC* reporter (*TOC1p* (TBSm):*LUC*) and transformed it into *Arabidopsis*. We found that these mutations significantly inhibited transcription of the reporter ($P < 0.001$), indicating that TGAs are transcription activators of *TOC1* (Fig. 2e). A direct role that NPR1/TGA plays in regulating *TOC1* expression was further confirmed through chromatin immunoprecipitation (ChIP) in which association of NPR1 to TBS in the *TOC1* promoter was significantly enhanced upon SA induction (Fig. 2f).

TOC1 is unlikely to be the only clock gene regulated by NPR1, because lowering the *TOC1* level shortens the clock period whereas elevating the level lengthens the period^{22,23}. However, no such perturbation was observed in *npr1* (Extended Data Fig. 3d) or after SA treatment (Fig. 1d, e). Moreover, SA treatment at dawn should have caused an immediate induction in *TOC1* expression instead of a 12-h delay (Fig. 1d). To systematically search for other NPR1-targeted clock genes, we performed mathematical modelling using the P2012 circadian model¹³ under the assumption that NPR1 is also a transcriptional activator of other clock genes (*X* and *Y* in Fig. 3a) (see Methods and Extended Data Figs 6 and 9 for details).

We first optimized the P2012 model to fit the *TOC1* expression in *npr1* (Fig. 2a), which was a single parameter fit (that is, basal expression in the absence of functional NPR1). The heat map of the best least-squares fit showed a characteristic ‘crosshair’ pattern centred on *PSEUDO-RESPONSE REGULATOR 7* (*PRR7*) (Fig. 3b), indicating that the basal regulation of *PRR7* by NPR1 best explains the unchanged

d, e, *TOC1p:LUC* activity rhythms in plants treated with water (CK) or SA at subjective dawn (**d**) and dusk (**e**) (mean \pm s.e.m.; $n = 6$). Arrows indicate treatment time; a.u., arbitrary unit. Bar graphs, mean \pm s.e.m. (Holm–Sidak test; ** $P < 0.01$; **** $P < 0.0001$).

TOC1 period in *npr1* (Extended Data Fig. 6a, b). This prediction was verified using qPCR in which *PRR7* transcript levels in *npr1* were found to be significantly lower than wild type (WT) (Fig. 3c). The second fitting for SA-induction data involved multiple parameters. We used our fixed basal expression parameter and NPR1 western data (Fig. 2b and Extended Data Fig. 4) to fit the *TOC1* expression from Fig. 2a. The resulting heat map showed a ‘crosshair’ pattern for *LHY/CCA1* (Fig. 3d), suggesting that either one or both of these genes is responsive to SA through the function of NPR1. Using qPCR, we found that while *CCA1* and *EARLY FLOWERING 3* (a negative control) did not respond to SA, the amplitude of *LHY* expression was significantly elevated by SA ($P < 0.05$) (Extended Data Fig. 6c–e) as predicted by our model. This result was further confirmed using the *LHYp:LUC* reporter (Fig. 3e). Consistently, the amplitude of basal *LHY* expression was reduced in *npr1* ($P < 0.05$) whereas that of *CCA1* remained unchanged (Extended Data Fig. 6f, g). Because *LHY* is an antagonist of *TOC1* in the clock, induction of *LHY* by SA explains the delayed increase in *TOC1* after SA treatment at dawn (Extended Data Fig. 6i, j) when *LHY* has its highest expression. This balanced network architecture of NPR1 regulating both the morning-phased *LHY* and the evening-phased *TOC1* (Fig. 3a) strengthens the clock when the redox rhythm is perturbed.

To investigate the effect of reinforced circadian clock on plant immunity, we examined SA-induced resistance against bacteria in a *toc1* mutant and found it to be more sensitive to induction than WT and *npr1* (Fig. 4a). While *TOC1* negatively regulates resistance against bacteria, *CCA1* and *LHY* have been reported to positively regulate resistance against bacteria and oomycetes^{24,25}, timing immunity for the morning when temperature and humidity are the most favourable for infection²⁵. We hypothesized that SA/NPR1-mediated induction of both morning and evening components of the circadian clock plays a

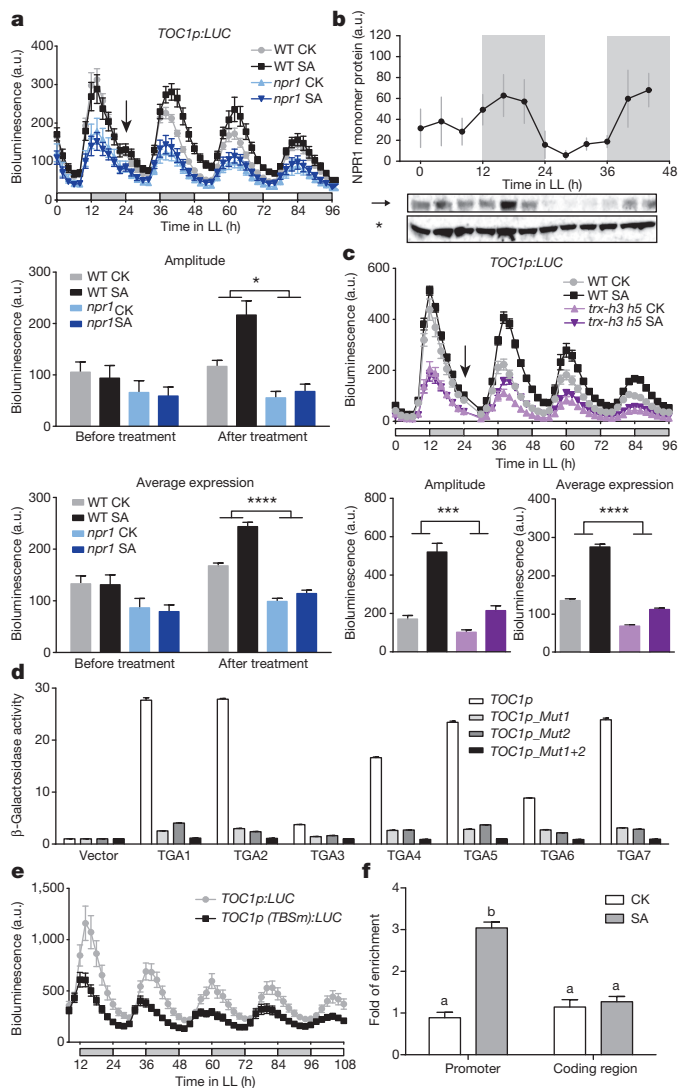


Figure 2 | SA-regulation of *TOC1* depends on nuclear NPR1. **a**, *TOC1p:LUC* activity rhythms in WT and *npr1-3* (**a**) or *trx-h3 trx-h5* (*trx-h3 h5*) (**c**) treated with water (CK) or 1 mM SA at subjective dawn (arrow) under constant light (mean \pm s.e.m.; $n = 6$). Bar graphs, mean \pm s.e.m. (two-way analysis of variance (ANOVA); * $P < 0.05$; *** $P < 0.001$; **** $P < 0.0001$). **b**, NPR1 monomer (arrow) quantified using the non-specific bands (*) as a loading control. Data are mean \pm s.e.m. ($n = 3$). The uncropped version is shown in Extended Data Fig. 4a. **d**, β -Galactosidase reporter activities shown as fold changes over the vector control. Mut1 and Mut2, mutants of two TGA-binding sites (TBSm). Data are mean \pm s.e.m. ($n = 3$). **e**, Luciferase activity rhythms of *TOC1p:LUC* and *TOC1p(TBSm):LUC* (mean \pm s.e.m., $n = 20$ T1-transformants). **f**, ChIP experiments were performed for the *TOC1* gene using 35S:NPR1-GFP (in *npr1-1*) plants. Data are mean \pm s.e.m. ($n = 3$; Tukey's multiple comparisons test; $P < 0.0001$).

major role in maintaining this diurnal difference in plant sensitivity to pathogen challenge.

To test this, we first examined induction of *WRKY40*, a direct target of *TOC1* (ref. 26), and *PR1*, a direct target of NPR1 (ref. 27), 3 h after SA application either in the subjective morning (ZT24) or evening (ZT36) under constant light. Both defence genes had higher induction after the morning treatment (Fig. 4b). We next performed microarray to investigate this time-of-day-specific sensitivity globally (GSE61059) (patterns of representative genes verified by qPCR shown in Extended Data Fig. 7a, b). We found more genes showing higher induction by morning SA treatment than the evening treatment (like *PR1*) (Fig. 4c). They were mainly defence-related genes (Fig. 4d). In contrast, a larger number of genes appeared to be more repressed after the evening

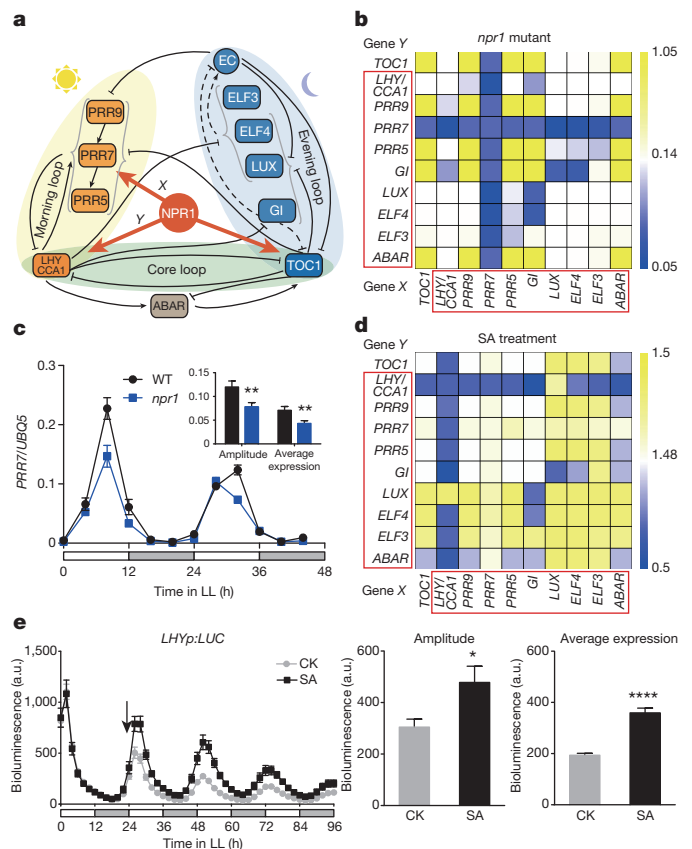


Figure 3 | NPR1 regulates transcription of multiple clock genes. **a**, NPR1 regulates transcription of genes in the P2012 version of *Arabidopsis* TTFL clock¹³. X, Y, query genes for mathematical modelling. Arrows, transcriptional activation. Blocked arrows, repression. Dashed lines, post-translational interactions or regulation. **b**, **d**, The least-squares fitting results of different query genes (X, Y) to the *npr1* data (**b**) and the SA-treated WT data (**d**) in Fig. 2a. The colour bars indicate the least-squares residual for each gene combination. Lower residual indicates a better fit. **c**, *PRR7* mRNA in WT and *npr1*. Data are mean \pm s.e.m. ($n = 3$; t -test; ** $P < 0.01$). **e**, *LHYp:LUC* activity rhythms in plants treated with water (CK) or SA at subjective dawn (arrow) (mean \pm s.e.m.; $n = 6$). Bar graphs, mean \pm s.e.m. (t -test; * $P < 0.05$; **** $P < 0.0001$).

treatment and they were enriched in plant growth and development (Fig. 4d). Furthermore, promoter analysis of the differentially induced genes in the morning showed significant enrichments for both *cis*-elements bound by CCA1/LHY and *TOC1*, and those of the differentially repressed genes in the evening had significantly enriched *cis*-elements bound by CCA1/LHY (Extended Data Fig. 7c, d). Collectively, these data strongly support our hypothesis that acute perturbation in redox rhythm caused by SA treatment leads to increased expression of both positive and negative regulators of defence, but with the former in the morning and the latter in the evening (Fig. 3a and Extended Data Fig. 7c, d). This may increase the diurnal differences in sensitivity to pathogen challenge in plants.

Gating defence towards the morning may also be a mechanism for plants to minimize interference on growth at night^{28,29}. We observed that SA treatment of dark-grown plants could help sustain a more robust circadian rhythm ($P < 10^{-15}$) than mock treatment ($P < 0.001$) (Fig. 4e), but induced a severe loss in fresh weight (Fig. 4f). This is consistent with our hypothesis that untimely induction of immunity at night is detrimental to plant growth. Besides gating immune response, reinforcement of the clock may also help increase photosynthesis and negate the redox perturbation through enhanced expression of *CHLOROPHYLL BINDING PROTEIN 2* (*CAB2*) and evening-phased *CATALASE 3* (*CAT3*) but not the morning-phased *CAT2* (Fig. 4g).

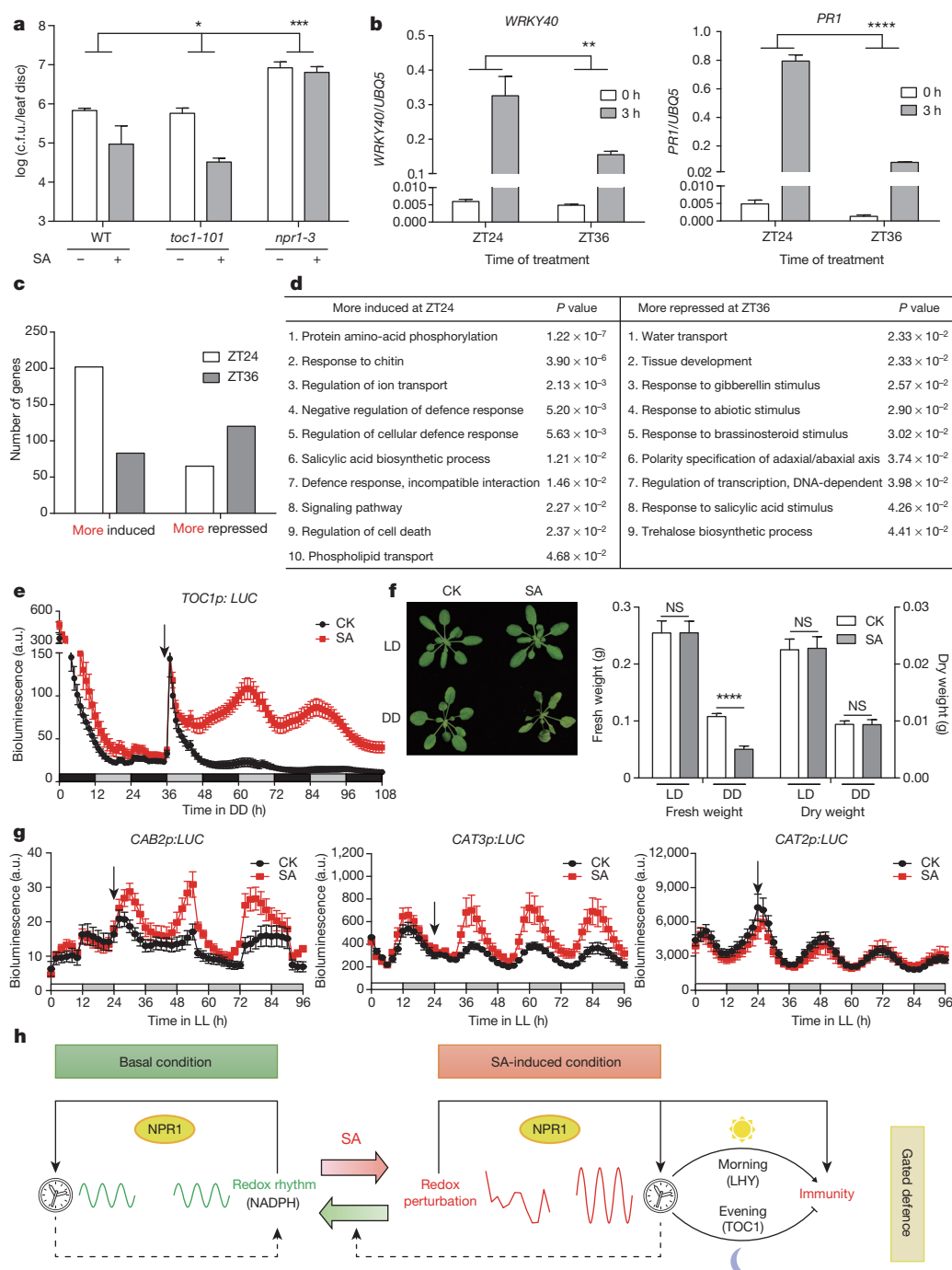


Figure 4 | SA reinforces the circadian clock to gate immune response.

a, *Pseudomonas syringae* pathovar *maculicola* ES4326 growth in plants pre-treated with water (–) or SA (+) (24 h) 3 days after infection; c.f.u., colony-forming units. Data are mean \pm 95% confidence intervals ($n = 8$; two-way ANOVA; * $P < 0.05$; *** $P < 0.001$). **b**, *WRKY40* and *PR1* expression in plants 0 or 3 h after SA treatment. Data are mean \pm s.e.m. ($n = 3$; two-way ANOVA; ** $P < 0.01$; **** $P < 0.0001$). **c**, Time-of-day-specific transcriptome changes in response to SA treatment. **d**, Enriched gene ontology categories.

e, **g**, *TOC1p:LUC* (**e**), *CAB2p:LUC*, *CAT3p:LUC*, and *CAT2p:LUC* (**g**) activity rhythms in water (CK)- or SA-treated plants (mean \pm s.e.m.; $n = 11$ in **e** and $n = 6$ in **g**). DD, constant dark. Darker and lighter bars represent subjective days and nights, respectively. **f**, Symptom (left) and fresh/dry weight (right) of plants treated with water (CK) or SA under constant dark (DD) or diurnal conditions (LD) (mean \pm s.e.m.; $n = 6$; t -test; **** $P < 0.0001$; NS, non-significant). **h**, A model showing the interactions between redox and the circadian clock in gating defence.

We propose that in *Arabidopsis* the daily redox rhythm is intrinsically linked with the basal expression of the circadian clock through NPR1 (Fig. 4h and Extended Data Fig. 8). Perturbation in redox rhythms caused by SA during pathogen challenge is sensed by NPR1 to trigger defence gene expression and to reinforce the circadian clock. The wiring of NPR1 to defence genes as well as to the clock shows how plants gate their immune responses towards the morning to anticipate infection while minimizing fitness costs on plant growth, which occurs mainly at night³⁰.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 April 2014; accepted 7 April 2015.

Published online 22 June 2015.

1. O'Neill, J. S. *et al.* Circadian rhythms persist without transcription in a eukaryote. *Nature* **469**, 554–558 (2011).

2. O'Neill, J. S. & Reddy, A. B. Circadian clocks in human red blood cells. *Nature* **469**, 498–503 (2011).
3. Edgar, R. S. *et al.* Peroxiredoxins are conserved markers of circadian rhythms. *Nature* **485**, 459–464 (2012).
4. Peek, C. B. *et al.* Circadian clock NAD⁺ cycle drives mitochondrial oxidative metabolism in mice. *Science* **342**, 1243417 (2013).
5. Ramsey, K. M. *et al.* Circadian clock feedback cycle through NAMPT-mediated NAD⁺ biosynthesis. *Science* **324**, 651–654 (2009).
6. O'Neill, J. S., Maywood, E. S., Chesham, J. E., Takahashi, J. S. & Hastings, M. H. cAMP-dependent signaling as a core component of the mammalian circadian pacemaker. *Science* **320**, 949–953 (2008).
7. Lai, A. G. *et al.* CIRCADIAN CLOCK-ASSOCIATED 1 regulates ROS homeostasis and oxidative stress responses. *Proc. Natl Acad. Sci. USA* **109**, 17129–17134 (2012).
8. McClung, C. R. The genetics of plant clocks. *Adv. Genet.* **74**, 105–139 (2011).
9. Nagel, D. H. & Kay, S. A. Complexity in the wiring and regulation of plant circadian networks. *Curr. Biol. CB* **22**, R648–R657 (2012).
10. Pokhilko, A. *et al.* The clock gene circuit in *Arabidopsis* includes a repressilator with additional feedback loops. *Mol. Syst. Biol.* **8**, 574 (2012).
11. Zhong, H. H. & McClung, C. R. The circadian clock gates expression of two *Arabidopsis* catalase genes to distinct and opposite circadian phases. *Mol. Gen. Genet.* **251**, 196–203 (1996).
12. Mou, Z., Fan, W. & Dong, X. Inducers of plant systemic acquired resistance regulate NPR1 function through redox changes. *Cell* **113**, 935–944 (2003).
13. Pokhilko, A., Mas, P. & Millar, A. J. Modelling the widespread effects of TOC1 signalling on the plant circadian clock and its outputs. *BMC Syst. Biol.* **7**, 23 (2013).
14. Kim, W. Y., Salome, P. A., Fujiwara, S., Somers, D. E. & McClung, C. R. Characterization of pseudo-response regulators in plants. *Methods Enzymol.* **471**, 357–378 (2010).
15. Goodspeed, D., Chehab, E. W., Min-Venditti, A., Braam, J. & Covington, M. F. *Arabidopsis* synchronizes jasmonate-mediated defense with insect circadian behavior. *Proc. Natl Acad. Sci. USA* **109**, 4674–4677 (2012).
16. Nawrath, C. & Metraux, J. P. Salicylic acid induction-deficient mutants of *Arabidopsis* express PR-2 and PR-5 and accumulate high levels of camalexin after pathogen inoculation. *Plant Cell* **11**, 1393–1404 (1999).
17. Fu, Z. Q. *et al.* NPR3 and NPR4 are receptors for the immune signal salicylic acid in plants. *Nature* **486**, 228–232 (2012).
18. Cao, H., Glazebrook, J., Clarke, J. D., Volk, S. & Dong, X. N. The *Arabidopsis* NPR1 gene that controls systemic acquired resistance encodes a novel protein containing ankyrin repeats. *Cell* **88**, 57–63 (1997).
19. Tada, Y. *et al.* Plant immunity requires conformational changes of NPR1 via S-nitrosylation and thioredoxins. *Science* **321**, 952–956 (2008).
20. Jahan, M. S. *et al.* Deficient glutathione in guard cells facilitates abscisic acid-induced stomatal closure but does not affect light-induced stomatal opening. *Biosci. Biotechnol. Biochem.* **72**, 2795–2798 (2008).
21. Despres, C. *et al.* The *Arabidopsis* NPR1 disease resistance protein is a novel cofactor that confers redox regulation of DNA binding activity to the basic domain/leucine zipper transcription factor TGA1. *Plant Cell* **15**, 2181–2191 (2003).
22. Mas, P., Alabadi, D., Yanovsky, M. J., Oyama, T. & Kay, S. A. Dual role of TOC1 in the control of circadian and photomorphogenic responses in *Arabidopsis*. *Plant Cell* **15**, 223–236 (2003).
23. Millar, A. J., Carre, I. A., Strayer, C. A., Chua, N. H. & Kay, S. A. Circadian clock mutants in *Arabidopsis* identified by luciferase imaging. *Science* **267**, 1161–1163 (1995).
24. Zhang, C. *et al.* Crosstalk between the circadian clock and innate immunity in *Arabidopsis*. *PLoS Pathog.* **9**, e1003370 (2013).
25. Wang, W. *et al.* Timing of plant immune responses by a central circadian regulator. *Nature* **470**, 110–114 (2011).
26. Huang, W. *et al.* Mapping the core of the *Arabidopsis* circadian clock defines the network structure of the oscillator. *Science* **336**, 75–79 (2012).
27. Wang, D., Weaver, N. D., Kesarwani, M. & Dong, X. Induction of protein secretory pathway is required for systemic acquired resistance. *Science* **308**, 1036–1040 (2005).
28. Shin, J., Heidrich, K., Sanchez-Villarreal, A., Parker, J. E. & Davis, S. J. TIME FOR COFFEE represses accumulation of the MYC2 transcription factor to provide time-of-day regulation of jasmonate signaling in *Arabidopsis*. *Plant Cell* **24**, 2470–2482 (2012).
29. Baldwin, I. T. & Meldau, S. Just in time: circadian defense patterns and the optimal defense hypothesis. *Plant Signal. Behav.* **8**, e24410 (2013).
30. Nozue, K. *et al.* Rhythmic growth explained by coincidence between internal and external cues. *Nature* **448**, 358–361 (2007).

Acknowledgements We thank R. McClung for sharing the *TOC1p:LUC*, *LHYp:LUC* and *CAB2p:LUC* transgenic lines; S. H. Wu for providing the *toc1-101* mutant; A. Millar for discussion on the project and advice on the modelling; S. Spoel for suggestions on an experiment; and P. Benfey for critiquing the manuscript. This work was supported by grants from the National Institutes of Health (NIH) (1R01-GM099839-01, 2R01-GM069594-09) and by the Howard Hughes Medical Institute and the Gordon and Betty Moore Foundation (through grant GBMF3032) to X.D., and a Defense Advanced Research Projects Agency (DARPA) Biochronicity Grant (DARPA-BAA-11-66), NIH Director's New Innovator Award (DP2 OD008654-01), and Burroughs Wellcome Fund CASI Award (BWF 1005769.01) to N.E.B.

Author Contributions M.Z., W.W., M.M., and J.M. performed the experiments and statistical analysis. S.K. and N.E.B. identified additional links of NPR1 to the circadian clock through mathematical modelling. X.D. supervised the project. M.Z., W.W., S.K., N.E.B., and X.D. wrote paper. All authors discussed the results and commented on the manuscript.

Author Information The microarray data have been deposited in Gene Expression Omnibus under accession number GSE61059. The computer code is available upon request. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to X.D. (xdong@duke.edu) or N.E.B. (nb69@duke.edu), for the mathematical modelling.

METHODS

No statistical methods were used to predetermine sample size.

Plant materials. The *TOC1p:LUC* (Col-0), *LHYp:LUC* (Col-0)¹⁴, and *CAB2p:LUC* (Col-0) seeds were provided by R. McClung and the *toc1-101* mutant³¹ by S.-H. Wu. Mutants of *npr1-3* (ref. 18), *sid2* (ref. 16), *trx-h3* (ref. 19), and *trx-h5* (ref. 19) were used to cross with the luciferase reporter lines. *35S:NPR1-GFP* (in *npr1-1*)¹² plants were used in ChIP experiments. To generate *CAT3p:LUC*, *CAT2p:LUC* homozygous lines and different T1 lines of *TOC1p:LUC* and *TOC1p(TBSm):LUC* (*TOC1* promoter with mutated TGA-binding sites), WT *CAT3*, *CAT2*, and *TOC1* promoters and mutated *TOC1* promoter (amplified using QuikChange Lighting Multi Site-directed mutagenesis kit, Agilent Technologies) were cloned into the pDONR207 vector (Invitrogen) through the Gateway BP reaction (Invitrogen) and then transferred to the destination vector pGWB235 (ref. 32) through the Gateway LR recombination reaction (Invitrogen). *Agrobacterium*-mediated transformation of *Arabidopsis* was performed as previously described using WT plants³³. Homozygous T3 lines of *CAT2p:LUC*, *CAT3p:LUC*, and different T1 lines of *TOC1p:LUC* and *TOC1p(TBSm):LUC* were selected and used for the luciferase imaging experiment. All primer sequences used for making the transgenic constructs are listed in Extended Data Table 1.

NADP⁺ and NADPH measurement. Three-week-old WT (Col-0) plants grown in soil under diurnal condition (12 h light/12 h dark) were treated with water or 1 mM SA at subjective dawn and samples were collected every 4 h for 2 days under constant light conditions. NADP⁺ and NADPH were measured according to Queval and Noctor³⁴ with modifications. Briefly, 50 mg of 3-week-old leaves were pulverized in liquid nitrogen using Genogrinder and extracted using 10 mM Tris-HCl (pH 8.0, 1 ml Tris-HCl per 100 mg tissue). The homogenate was centrifuged at 16,000g for 10 min at 4 °C. The supernatant was separated into two 0.2 ml aliquots. To extract NADP⁺, 50 µl 1 M HCl was added to one 0.2 ml aliquot. The mixture was heated in boiling water for 1 min. Then 25 µl MES (pH 5.6) was added and the pH of the extract was adjusted to 5–6 using 0.2 M NaOH. To extract NADPH, 50 µl 1 M NaOH was added to the other 0.2 ml aliquot. The mixture was heated in boiling water for 1 min. Then 25 µl MES (pH 5.6) was added and the pH of the extract was adjusted to 7–8 using 0.2 M HCl. Three 20 µl aliquots of the NADP⁺ and the NADPH extracts were used as technical replicates. Samples containing only the extraction buffer were used as blank. The measurement of the samples and the derivation of the standard curves were performed according to Queval and Noctor³⁴.

RNA extraction and qPCR. Three-week-old plants grown in soil under diurnal conditions (12 h light/12 h dark) were treated with water or 1 mM SA at subjective dawn and samples were collected every 4 h for 2 days under constant light conditions. RNA extraction was performed as previously described¹⁸. Complementary DNA synthesis (SuperScript III, Invitrogen) and qPCR (SYBR Green, Roche) were performed according to the manufacturer's protocols. All primer sequences used for qPCR are listed in Extended Data Table 1.

Luciferase activity measurement. Plants grown in soil with under 12 h light/12 h dark cycles for 3 weeks were sprayed with 2.5 mM luciferin (Gold Biotechnology) in 0.02% Triton X-100 (Sigma) to activate and deplete pre-existing luciferase because of its instability in the presence of the substrate. Then the plants were transferred into constant light condition. SA (1 mM; Sigma) or water (as control) was sprayed 24 h later (ZT24). The fifth and sixth leaves were harvested and rinsed three times in 50 ml water. Luciferase was extracted and relative activity was measured according to the manufacturer's protocol (Luciferase Assay System, Promega).

Luciferase imaging. Plants grown in soil under 12 h light/12 h dark cycles for 3 weeks were sprayed with 2.5 mM luciferin (Gold Biotechnology) in 0.02% Triton X-100 (Sigma) 1 day before luciferase imaging. Plants were then placed into the imaging system (Nightshade LB985) under either constant light or dark conditions and assayed for bioluminescence by acquiring images with exposure time of 20 min. To test the effect of SA or GSHmee, 1 mM SA (Sigma)/3 mM GSHmee (Sigma) or water (as control) was sprayed at different indicated times. Subsequent quantifications of bioluminescence intensity were performed using Image J.

Analysis of circadian rhythms. The quantified time-course bioluminescence data were decomposed into a line and a sine wave with exponentially decaying

$$\text{amplitude } Y = \text{amplitude} \times e^{-Kt} \times \sin\left(\frac{2\pi t}{\text{period}} + \text{phase shift}\right) + a \times t + b$$

using GraphPad Prism 6. The intercept of the line at the y axis ('b') was considered as the average expression level. The period and amplitude were inferred from the sine wave. The exponential decay was used to account for the dampening of bioluminescence over time. The best-fitted value, standard error, and degrees of freedom were used for statistical analysis.

Western blot. Three-week-old WT (Col-0) plants grown in soil under diurnal condition (12 h light/12 h dark) were treated with water or 1 mM SA at subjective dawn and samples were collected every 4 h for 2 days under constant light. Detection of the NPR1 monomer protein on a non-reducing SDS–polyacrylamide gel electrophoresis (SDS–PAGE) gel was performed as previously described using an antibody against NPR1 (ref. 12).

Yeast one-hybrid assay. The *TOC1* promoter was first cloned into the pDONR P4-P1R vector (Invitrogen) through the Gateway BP reaction. The entry clones were recombined into destination vectors pMW#2 (Invitrogen) and pMW#3 (Invitrogen). Mutagenesis of the *TOC1* promoter was performed using a QuikChange Lighting Multi Site-directed mutagenesis kit (Agilent Technologies) according to the instruction manual. *TOC1p_Mut1* (the *TOC1* promoter mutated in the first TGA-binding site), *TOC1p_Mut2* (the *TOC1* promoter mutated in the second TGA-binding site), and *TOC1p_Mut1+2* (the *TOC1* promoter mutated in both TGA-binding sites) were cloned into destination vectors pMW#2 and pMW#3 through the Gateway cloning kit (Invitrogen). The coding sequences of TGAs were cloned into pDONR207 and subsequently transferred into pDEST-AD by the Gateway LR reactions. Transformation of constructs into the yeast strain YM4271 was performed as previously described³⁵. β-Galactosidase reporter activities were measured using ONPG as the substrate³⁶ and normalized to the control with an empty vector pDEST-AD. All primer sequences used for yeast one-hybrid assay (Y1H) are listed in Extended Data Table 1.

ChIP. Three-week-old soil-grown *35S:NPR1-GFP* (in *npr1-1*)¹² plants were treated with either water (CK) or 1 mM SA at dusk and samples were collected 3 h after treatment. ChIP was performed as described previously³⁷. Immunoprecipitation was performed using a polyclonal antibody against GFP (Ab290, Abcam) and Dynabeads Protein G (Invitrogen). The purified ChIP samples were subject to qPCR using primer pairs for the promoter region (−639 to −589 base pairs (bp) upstream of the start codon) and the coding region (+753 to +803 bp downstream of the start codon) of *TOC1*. Fold of enrichment was calculated using the comparative C_T method³⁸ using the input samples as normalizers. All primer sequences used for ChIP are listed in Extended Data Table 1.

The mathematical model of the *Arabidopsis* circadian clock. We applied the P2012 plant circadian model from the Plant Systems Modelling portal¹³ to elucidate new connections between SA signalling through NPR1 and known plant circadian genes. This numerical ordinary differential equation model in MATLAB consists of 32 ordinary differential equations and includes transcription terms for ten genes, which are *LHY/CCA1*, *PSEUDO-RESPONSE REGULATOR 9*, *7*, *5* (*PRR9*, *PRR7*, *PRR5*), *TOC1*, *EARLY FLOWERING 4*, *3* (*ELF4*, *ELF3*), *LUX ARRHYTHMO* (*LUX*), *GIGANTEA* (*GI*) and *ABA receptor* (*ABAR*). The 133 parameters in P2012 were previously fitted to multiple data sets in various light–dark photoperiods, different genetic backgrounds, and ABA signalling. It is important to note that the P2012 model was designed to understand and predict changes in period and phase when perturbed by genetic or environmental variations. The model does not aim to reflect the exact transcriptional profiles or the absolute protein concentrations.

It was recently shown that plant circadian models exhibit a 'period overshoot' when transitioning from LD to LL cycles³⁹. This period overshoot introduced a constant phase delay in the LL data relative to the LD data. Since this phenomenon was not observed in the experimental data, it is an artefact of the mathematical model. For example, the P2012 model predicts that *TOC1* mRNA peaks at ZT18 (that is, 6 h after subjective light-to-dark transition) under LL conditions¹³ instead of the real peak time at ZT12. Moreover, our experiments indirectly measured *TOC1* expression via the luciferase reporter, which is known to exhibit delays^{40,41}. This delay was deduced to be 2 h because the luciferase reporter peaked at ZT14 (Fig. 1d) whereas the *TOC1* mRNA peaked at ZT12 (Fig. 1c).

To take this 2-h delay and the 'period overshoot' in the model into consideration, we empirically measured a 4-h delay between simulation *TOC1* mRNA levels and our luminescence data. This total 4-h delay was inferred by aligning the second peak after the LD to LL transition in our luciferase experiments (38 h) and *TOC1* mRNA in the model (42 h). We subsequently used this 4-h delay to correctly align and fit the P2012 simulation *TOC1* mRNA to our experimental luciferase data.

Addition of NPR1 regulation to the circadian clock model. While keeping the original P2012 parameters fixed, we added NPR1 as a transcriptional activator of *TOC1*, as it has been shown experimentally. We also added NPR1 as a transcriptional activator of two additional clock genes ('query pair'). Our goal was to systematically determine which query pair best fitted our measured *TOC1p:LUC* expression in WT and *npr1-3* in mock- and SA-treated plants. We multiplied the P2012 transcriptional synthesis term of *TOC1* and each gene in the query pair by

their own NPR1-dependent regulatory function $F(t)$ (that is, non-competitive activation). Each regulatory function F has the form

$$F(t) = n_b + n_a \frac{[\text{NPR1}(t)]}{[\text{NPR1}(t)] + K_d}$$

where $[\text{NPR1}(t)]$ is the NPR1 monomer concentration over time, n_b is the basal, NPR1-independent transcription level of the gene of interest, n_a is the maximum NPR1-activated transcription level of the gene, and K_d is the effective DNA-binding dissociation constant for the gene. The $[\text{NPR1}(t)]$ monomer levels for mock-treated and SA-treated plants were taken from western blot data in Fig. 2b and Extended Data Fig. 4. The NPR1 data for the mock-treated and SA-treated plants were then averaged, normalized, and linearly interpolated to serve as an input function for modelling (Extended Data Fig. 9a, b).

Least-squares fitting of the *TOC1p:LUC* data. For every query pair and the *TOC1* gene, we optimized n_b , n_a , K_d parameters (among the nine parameters, six are independent; see below) to give the best least-squares fit of the *TOC1* mRNA in the model to the patterns of *TOC1p:LUC* expression in WT and *npr1-3* with mock- or SA-treated plants over several circadian cycles (Fig. 2a).

Because the time of sampling and the waveforms were different between our experiments and the P2012 model, our luciferase data could not be fitted directly to the model. To solve the sampling time discrepancy, data points from our experiments and the P2012 data sets were interpolated (via cubic spline) to a time resolution of 0.1 h. To circumvent the waveform issue, we first calculated the ratio (R) of *npr1-3*/WT and SA-treated/mock-treated WT in experimental *TOC1p:LUC* data:

$$R_{\text{npr1}}(t) = \frac{\text{TOC1}_{\text{npr1}}(t)}{\text{TOC1}_{\text{WT}}(t)}, \quad R_{\text{SA}}(t) = \frac{\text{TOC1}_{\text{SA}}(t)}{\text{TOC1}_{\text{WT}}(t)}$$

where TOC1_{WT} , $\text{TOC1}_{\text{npr1}}$ and TOC1_{SA} are interpolated experimental data for *TOC1p:LUC* in (1) mock-treated WT, (2) mock-treated *npr1* mutant, and (3) SA-treated WT, respectively. We then created target P2012 simulation of $\text{TOC1}_{\text{npr1}}$ or TOC1_{SA} mRNA data by multiplying the TOC1_{WT} mRNA data in the simulation by R_{npr1} or R_{SA} . We optimized the regulatory function parameters (n_b , n_a , K_d) for each combination of query genes and *TOC1*. The parameter optimization used nonlinear least-squares fitting to minimize the sum of squared residual of *TOC1* expression (that is, squared difference between the model and the target *TOC1* mRNA profile). To account for the 4-h delay inherent to 'period-overshoot' in the P2012 model and the use of the reporter, we started fitting at 28 h, which corresponds to 24 h in our experiments. We fitted 3-day-long *npr1-3* mutant data and 2-day-long WT and SA treatment data.

Fitting of the *npr1-3* mutant data. For the first part of the fitting process, we set $[\text{NPR1}] = 0$, which resulted in a single parameter (n_b) fit for each candidate gene. Because we coupled NPR1-activation to *TOC1* and two other query genes, three parameters in total were fitted for each query pair. The optimal parameters n_b^* were restricted to a value between 0 and 1, where 0 represents no transcription in absence of NPR1 and 1 represents the absence of any regulation by NPR1. We used the function `fmincon` in MATLAB (2013b, MathWorks) with sequential programming algorithm without restrictions. We found the three parameters that minimized the least-squares residual of *TOC1* model output to target P2012 $\text{TOC1}_{\text{npr1}}$ data. Because nonlinear least-squares fitting uses a deterministic algorithm that can become trapped in a local minimum, we ran the simulations from 15 different random starting points to find the global minimum. We confirmed that nonlinear least-squares fitting of most query pairs converged to the same global minimum when started from random parameters (Extended Data Fig. 9c, d).

A plot of the best least-squares fit for each query gene showed a cross-hair pattern centred on *PRR7* in Fig. 3b. The *npr1* data (reduced *TOC1* amplitude, no change in phase or period) are mostly explained by including *PRR7* as an additional target of NPR1 regulation (Extended Data Fig. 6a, b), such that *PRR7* levels are reduced in the *npr1* mutant (Fig. 3c).

A role for NPR1 in regulating basal expression of *PRR7* is consistent with previous genetic data showing that lowering *TOC1* expression shortens the circadian period²³, whereas lowering *PRR7* expression lengthens the circadian period⁴². Analytical work has previously shown how mutations in opposing components in the clock can lead to unchanged period⁴³. We used our best-fit P2012 model to verify that lowered *PRR7* expression lengthens the period by ~2 h, whereas lowered *TOC1* expression shortens the period by ~2 h in *npr1*. Thus, the simultaneous, balanced reduction in expression of two opposing nodes (that is, *TOC1* and *PRR7*) explains why the period of *TOC1p:LUC* expression is not altered in *npr1*. **Mock-treated WT data constrain the parameters.** During our fitting procedure for mock-treated WT data, we discovered that optimal n_a^*, K_d^* always exhibited a simple relationship (Extended Data Fig. 9e). The constraint that explains this empirical relationship is that our final choice of n_a^*, K_d^* should not alter clock expression in mock-treated WT. Mathematically, n_a^*, K_d^* should have no effect,

on average, on gene expression in mock-treated WT. This condition restricts n_a^* to depend on K_d^* because the time-averaged $\langle [\text{NPR1}] \rangle$ in WT is normalized to 1, such that

$$\langle F \rangle \approx n_b^* + n_a^* \frac{\langle [\text{NPR1}] \rangle}{\langle [\text{NPR1}] \rangle + K_d^*} = n_b^* + n_a^* \frac{1}{1 + K_d^*} = 1.$$

Thus, we recovered the simple empirical relationship observed in Extended Data Fig. 9e as

$$n_a^* = (1 - n_b^*)(1 + K_d^*).$$

Fitting of the SA treatment data. By using the best-fit value for n_b^* from step 1 and the constraint from step 2 to fix n_a^* , we only needed to fit a single parameter (K_d). We restricted K_d to lie between 0 (that is, always maximum expression, insensitive to SA treatment and NPR1 monomer levels) and 5. We verified that allowing K_d to be as high as 50 did not significantly improve the least-squares fit. For each query pair, we ran nonlinear least-squares fit from 15 random starting points. We found that again, in general, they converged to the same global optimum (Extended Data Fig. 9f, g). Figure 3d shows a cross-hair pattern centred on *LHY/CCA1*, suggesting that *LHY/CCA1* activation by the induced, arrhythmic NPR1 levels during SA treatment should counteract the effect of *TOC1* induction. We noted that the best-fit solution from a *TOC1*-only case shows immediate *TOC1* induction after SA treatment (Extended Data Fig. 6i) instead of the observed delay until dusk. However, if we added *LHY* as an NPR1 target in addition to *TOC1*, the modelling results best fit our experimental data (Extended Data Fig. 6j). Consistently, *LHY* expression was found to be induced by SA (Fig. 3e), but reduced in the *npr1* mutant (Extended Data Fig. 6g).

Limitations of the model. There are limitations to our model. First, our model was only fitted to the expression of one gene (*TOC1p:LUC*) under three conditions (that is, WT with and without SA treatment, and *npr1*). Second, our model combines *LHY/CCA1* into one gene and cannot resolve the experimental differences that we observed in those genes (Fig. 3e and Extended Data Fig. 6c, d). Third, our model for the SA-induction data pre-sets *PRR7* at maximum expression even without SA treatment (that is, $K_d = 0$) (Extended Data Fig. 9g). This is unlikely to be an accurate reflection of a real physiological state. Last, all our experiments and modelling were done under constant light conditions and the additive (acute) light activation terms are effectively 0 because the hypothetical protein responsible for light activation has decayed to 0. These additive light terms should not be affected by our assumption of non-competitive activation by NPR1. Uncovering the proper relationship between the light-dependent terms and the NPR1-dependent terms (that is, competitive versus non-competitive activation) would require experiments under diurnal conditions. This is outside of the scope of the current paper.

Even though the modelling approach correctly predicted *PRR7* and *LHY* as direct targets of NPR1, it had mixed results with *ELF3*, which is not a direct target. Our model predicted *ELF3* to decrease after SA treatment and to increase in the *npr1* mutant. The lack of induction in *ELF3* by SA (Extended Data Fig. 6e) was consistent with the model prediction. However, the significantly decreased expression of *ELF3* observed in *npr1* (Extended Data Fig. 6h) was not in agreement with the model. This discrepancy suggests there are other links between NPR1 and the circadian clock that the current model cannot capture. A future model should be fitted to *ELF3p:LUC*, which would be an informative constraint.

Code availability. The MATLAB code, which was used to fit the modified P2012 plant circadian clock model to our *TOC1* luciferase data, is available upon request. A final SBML version of the modified P2012 model with our best-fit parameters can be downloaded from the BioModels Database (MODEL1506010000).

Bacterial infection. Three-week-old plants grown in soil were pre-treated with water or 1 mM SA at ZT12 and infiltrated 24 h later with *Pseudomonas syringae* pathovar *maculicola* (*Psm*) ES4326 (absorbance $A_{600 \text{ nm}} = 0.001$) as previously described¹². Briefly, eight plants per genotype per treatment were inoculated with *Psm* ES4326 and bacterial growth was measured 3 days after inoculation.

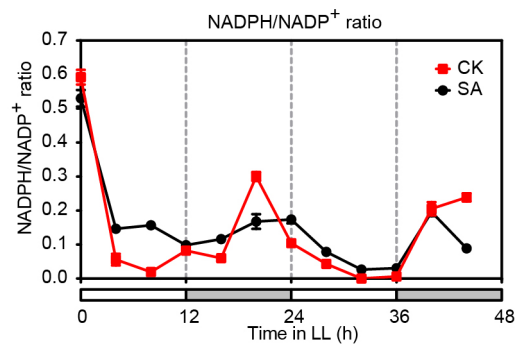
Microarray analysis. To test time-of-day-specific sensitivity to SA, 3-week-old soil-grown plants were transferred to constant light condition 1 day before treatment. Water or 1 mM SA was applied in the subjective morning (ZT24) or evening (ZT36). The fifth and sixth leaves were sampled 0 and 3 h after treatment. RNA was extracted, amplified, labelled, and hybridized to ATH1 GeneChip (Affymetrix) as previously reported²⁵. The arrays were normalized with RMA algorithm and centred to median. Two-way ANOVA ($P < 0.05$) and Student's *t*-test with multiple comparison correction ($P < 0.05$, fold changes > 2) were used to identify genes that were significantly more induced or more repressed by water or SA when treated at ZT24 or at ZT36. Athena program (<http://www.bioinformatics2.wsu.edu/Athena>) was used to identify *cis*-elements bound by CCA1/LHY including evening element, CCA1-binding site, CCA1 motif1 BS in CAB1, and CCA1 motif2 BS in CAB1. Hypergeometric distribution was used to determine

statistical significance. Enriched gene ontology categories were identified using BiNGO (<http://www.psb.ugent.be/cbd/papers/BiNGO/Home.html>).

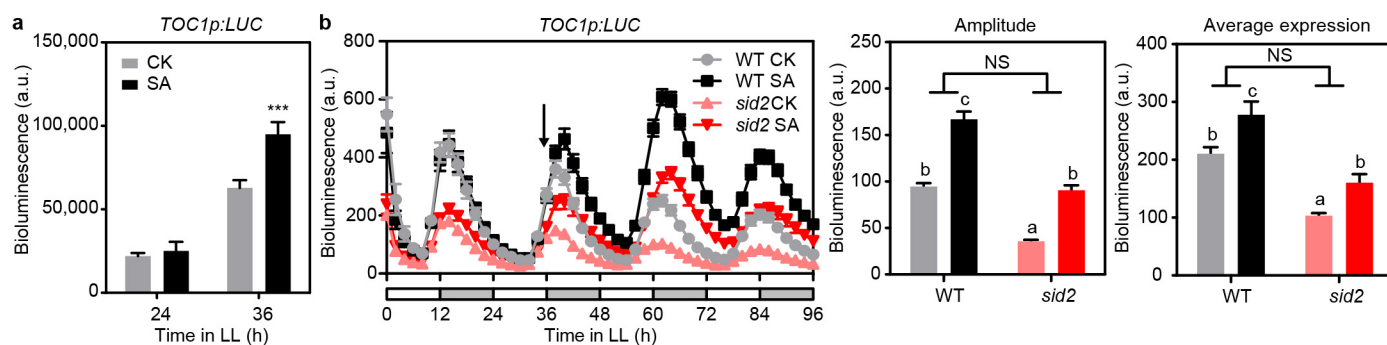
Fresh/dry weight measurement. Three-week-old soil-grown WT (Col-0) plants were transferred into constant dark condition or normal diurnal condition at dusk. After 36 h, water (control) or 1 mM SA was applied. Two days after treatment, pictures were taken and fresh/dry weight was measured.

Statistical analysis. Statistical analysis used GraphPad Prism 6 (GraphPad Software). All the centre values shown in the figures are means of technical (Figs 1a, b and 2d and Extended Data Fig. 1) or biological (all other figures where applicable) replicates. Experiments were repeated twice for Fig. 1a, b and *CAB2p:LUC* in Fig. 4g. All other experiments were repeated three times where applicable. Harmonic regression ($Y = a \sin(\pi t/12) + b \cos(\pi t/12) + c$) followed by ANOVA test was used to identify statistically significant oscillation. The null hypothesis was that all data across different time points were sampled from the same normal distribution. Student's *t*-test with multiple comparison correction was performed to identify statistically significant differences between mock and treated samples. Two-way ANOVA was used to assess significant interactions between genotype and treatment or between time of treatment and treatment. Significant interactions suggested the effect of the treatment was dependent on genotype or time of treatment. Tukey's multiple comparisons test was performed to identify the orders of samples that were significantly different from each other. All statistical tests were two-sided tests where applicable.

31. Wang, Y. *et al.* LIGHT-REGULATED WD1 and PSEUDO-RESPONSE REGULATOR9 form a positive feedback regulatory loop in the *Arabidopsis* circadian clock. *Plant Cell* **23**, 486–498 (2011).
32. Nakagawa, T. *et al.* Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J. Biosci. Bioeng.* **104**, 34–41 (2007).
33. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
34. Queval, G. & Noctor, G. A plate reader method for the measurement of NAD, NADP, glutathione, and ascorbate in tissue extracts: Application to redox profiling during *Arabidopsis* rosette development. *Anal. Biochem.* **363**, 58–69 (2007).
35. Deplancke, B., Dupuy, D., Vidal, M. & Walhout, A. J. A gateway-compatible yeast one-hybrid system. *Genome Res.* **14**, 2093–2101 (2004).
36. Pruneda-Paz, J. L., Breton, G., Para, A. & Kay, S. A. A functional genomics approach reveals CHE as a component of the *Arabidopsis* circadian clock. *Science* **323**, 1481–1485 (2009).
37. Gendrel, A. V., Lippman, Z., Yordan, C., Colot, V. & Martienssen, R. A. Dependence of heterochromatic histone H3 methylation patterns on the *Arabidopsis* gene DDM1. *Science* **297**, 1871–1873 (2002).
38. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C_T method. *Nature Protocols* **3**, 1101–1108 (2008).
39. Dodd, A. N., Dalchau, N., Gardner, M. J., Baek, S. J. & Webb, A. A. The circadian clock has transient plasticity of period and is required for timing of nocturnal processes in *Arabidopsis*. *New Phytol.* **201**, 168–179 (2014).
40. Finkenzadt, B. *et al.* Reconstruction of transcriptional dynamics from gene reporter data using differential equations. *Bioinformatics* **24**, 2901–2907 (2008).
41. Edwards, K. D. *et al.* Quantitative analysis of regulatory flexibility under changing environmental conditions. *Mol. Syst. Biol.* **6**, 424 (2010).
42. Farre, E. M., Harmer, S. L., Harmon, F. G., Yanovsky, M. J. & Kay, S. A. Overlapping and distinct roles of PRR7 and PRR9 in the *Arabidopsis* circadian clock. *Curr. Biol.* **15**, 47–54 (2005).
43. Rand, D. A., Shulgin, B. V., Salazar, D. & Millar, A. J. Design principles underlying circadian clocks. *J. R. Soc. Interface* **1**, 119–130 (2004).

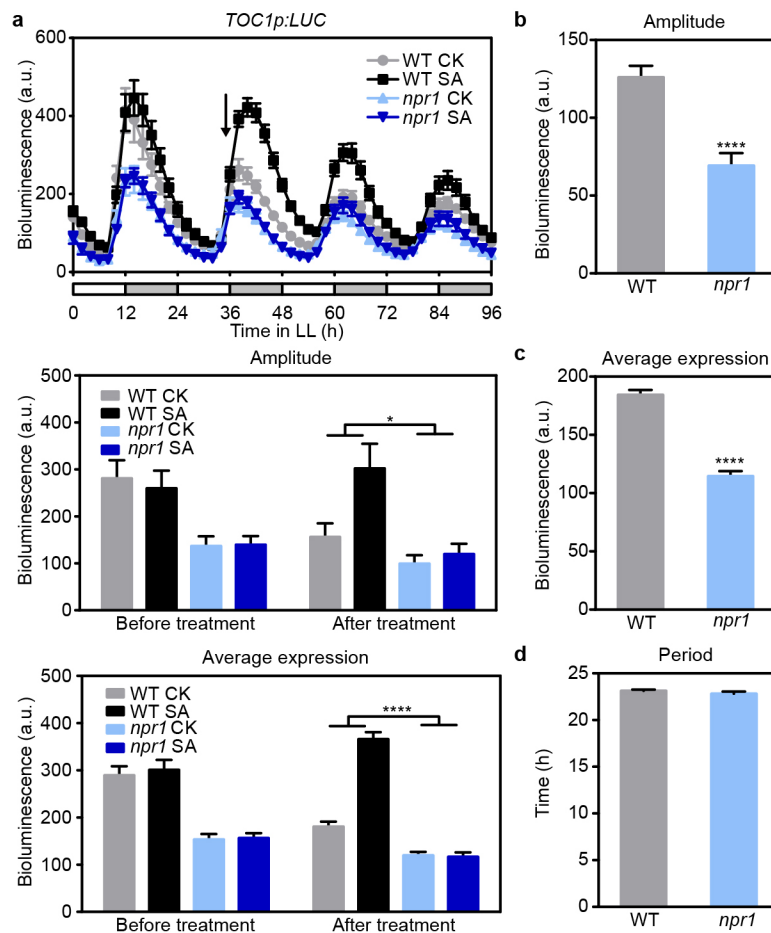


Extended Data Figure 1 | Circadian oscillation of the NADPH/NADP⁺ ratio. NADPH/NADP⁺ ratios in 3-week-old soil-grown plants derived from Fig. 1a, b. Water (CK) or 1 mM SA was applied at 0 h. Data are mean \pm s.e.m. ($n = 3$). White bars represent subjective days and grey bars represent subjective nights. Harmonic regression analysis suggests significant circadian oscillation of water-treated NADPH/NADP⁺ ratio ($P < 0.0001$).



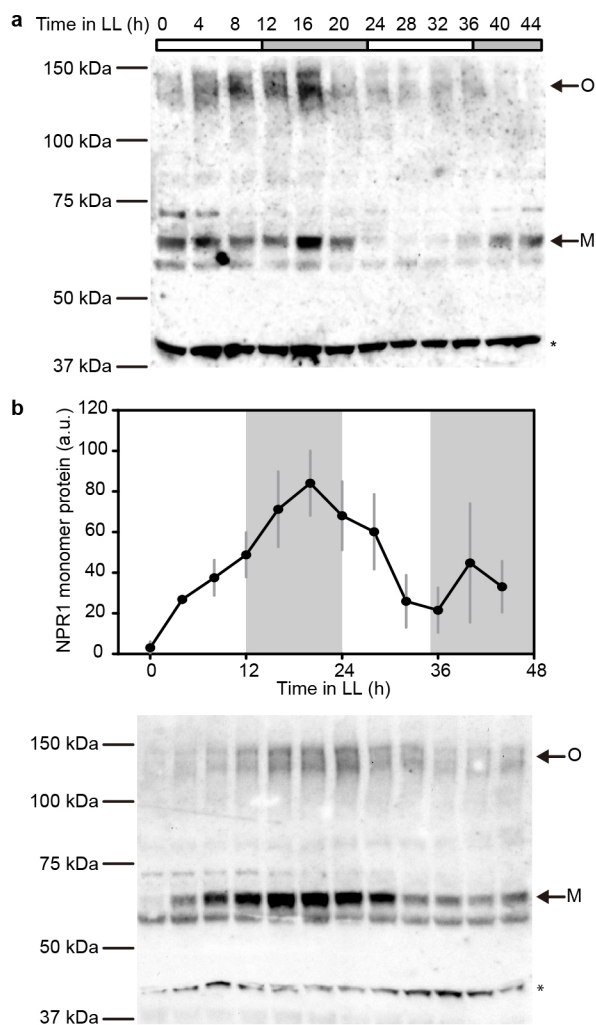
Extended Data Figure 2 | The effects of exogenous and endogenous SA on *TOC1* expression. **a**, Luciferase activity measurements using the *TOC1p:LUC* plant extracts. Relative luciferase activity of the fifth and sixth leaves from 3-week-old soil-grown *TOC1p:LUC* plants. Water (CK) or 1 mM SA was applied at ZT24. LL, constant light. a.u., arbitrary unit. Data are mean \pm s.e.m. ($n = 6$ biological replicates; t -test; *** $P < 0.001$). **b**, *TOC1p:LUC* activity rhythms in 3-week-old soil-grown WT and *sid2* plants treated with water (CK) or 1 mM SA at subjective dusk (black arrow) (mean \pm s.e.m., $n = 8$ plants).

White bars represent subjective days and grey bars represent subjective nights. The bar graphs represent the estimates of amplitude and average expression of *TOC1p:LUC*, respectively (mean \pm s.e.m.). The letters above the bars indicate statistically significant differences between groups at $P < 0.05$ (Tukey's multiple comparisons test). NS, non-significant (two-way ANOVA, non-significant interaction between genotype and treatment). This experiment was repeated three times with similar results.

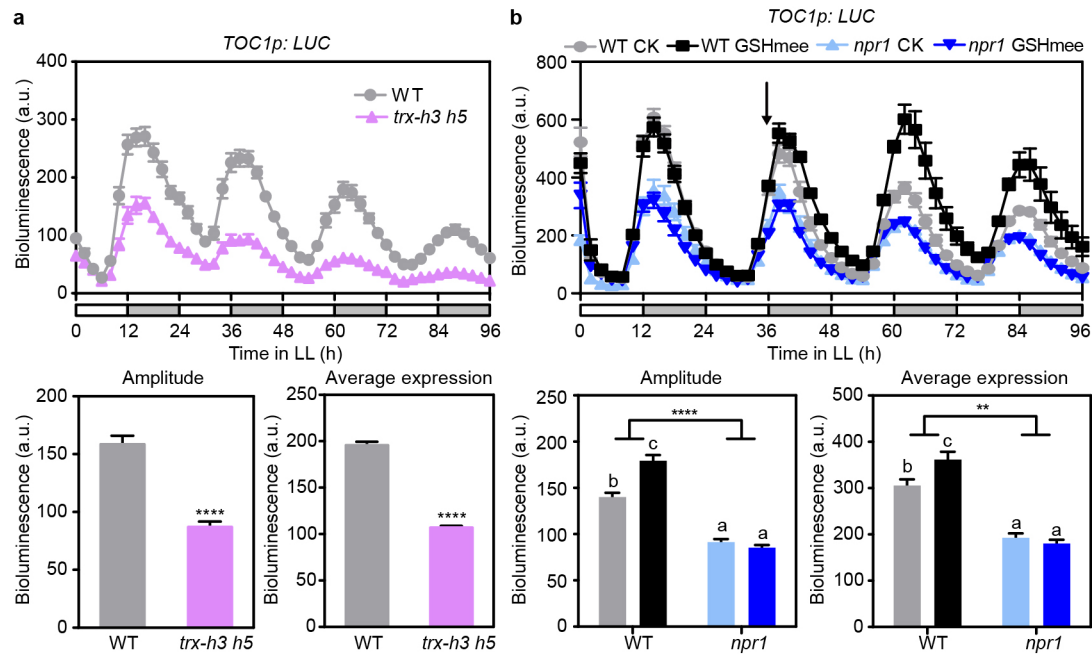


Extended Data Figure 3 | NPR1 regulates the amplitude and average expression of *TOC1p:LUC*. **a**, *TOC1p:LUC* activity rhythms in 3-week-old soil-grown WT and *npr1-3* plants treated with water (CK) or 1 mM SA at subjective dusk (black arrow) (mean \pm s.e.m.; $n = 6$ plants). LL, constant light. a.u., arbitrary unit. White bars represent subjective days and grey bars represent subjective nights. The bar graphs show the estimates of amplitude and

average expression level (mean \pm s.e.m.; two-way ANOVA; * $P < 0.05$; **** $P < 0.0001$). **b–d**, Estimates of amplitude (**b**), average expression (**c**), and period (**d**) of *TOC1p:LUC* in WT and *npr1-3*. Data are mean \pm s.e.m. (t -test; **** $P < 0.0001$). These experiments were repeated three times with similar results.

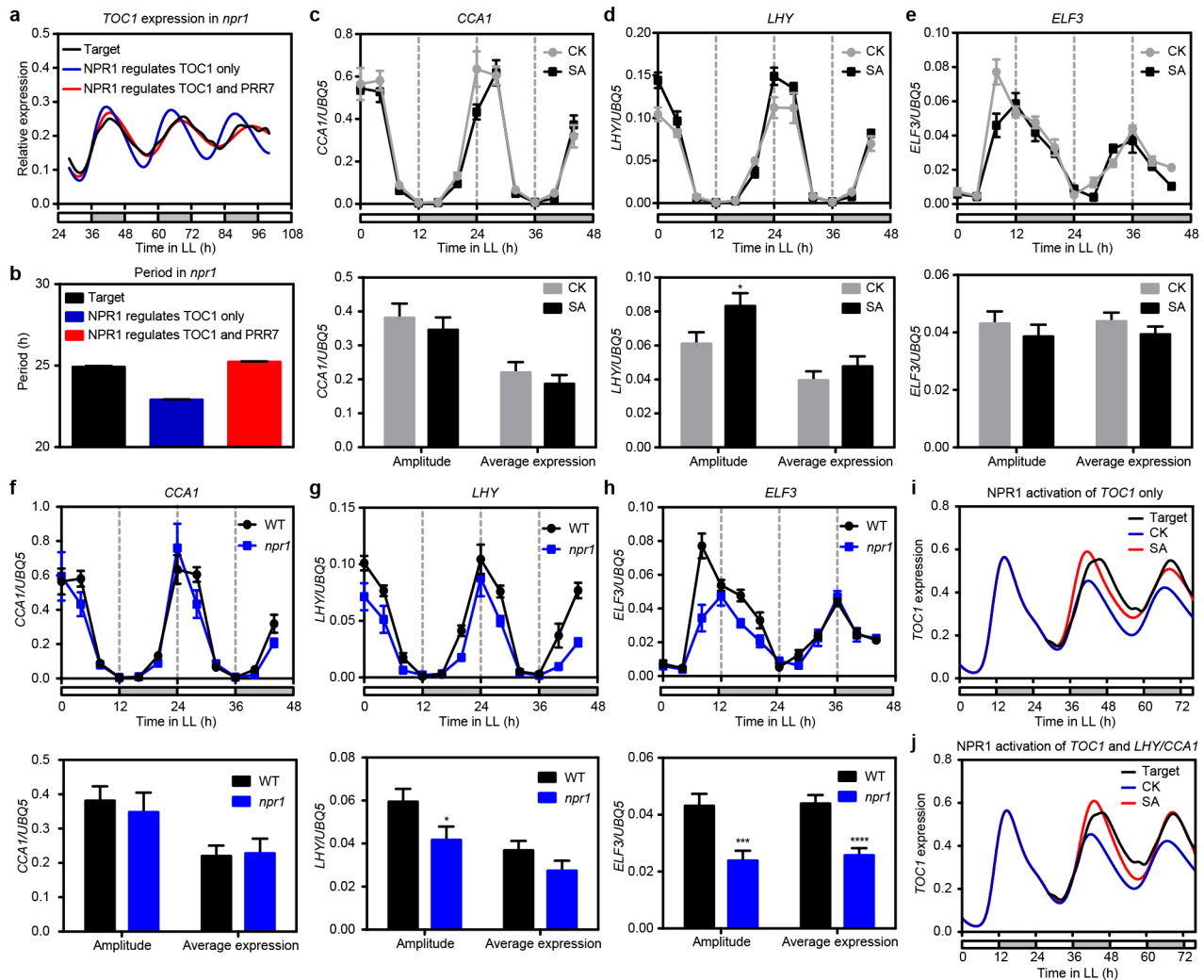


Extended Data Figure 4 | The abundance of NPR1 monomer under constant light conditions. NPR1 monomer (M) abundance in 3-week-old soil-grown plants without treatment (**a**; uncropped version of Fig. 2b) and after 1 mM SA treatment at 0 h (**b**) under constant light (LL) conditions. NPR1 protein were detected using western blot after non-reducing SDS-PAGE (**a**, **b**). NPR1 monomer protein was quantified using the non-specific band (*) as a loading control (**b**; mean \pm s.e.m.; $n = 3$ biological replicates). O, NPR1 oligomer. White bars represent subjective days and grey bars represent subjective nights.



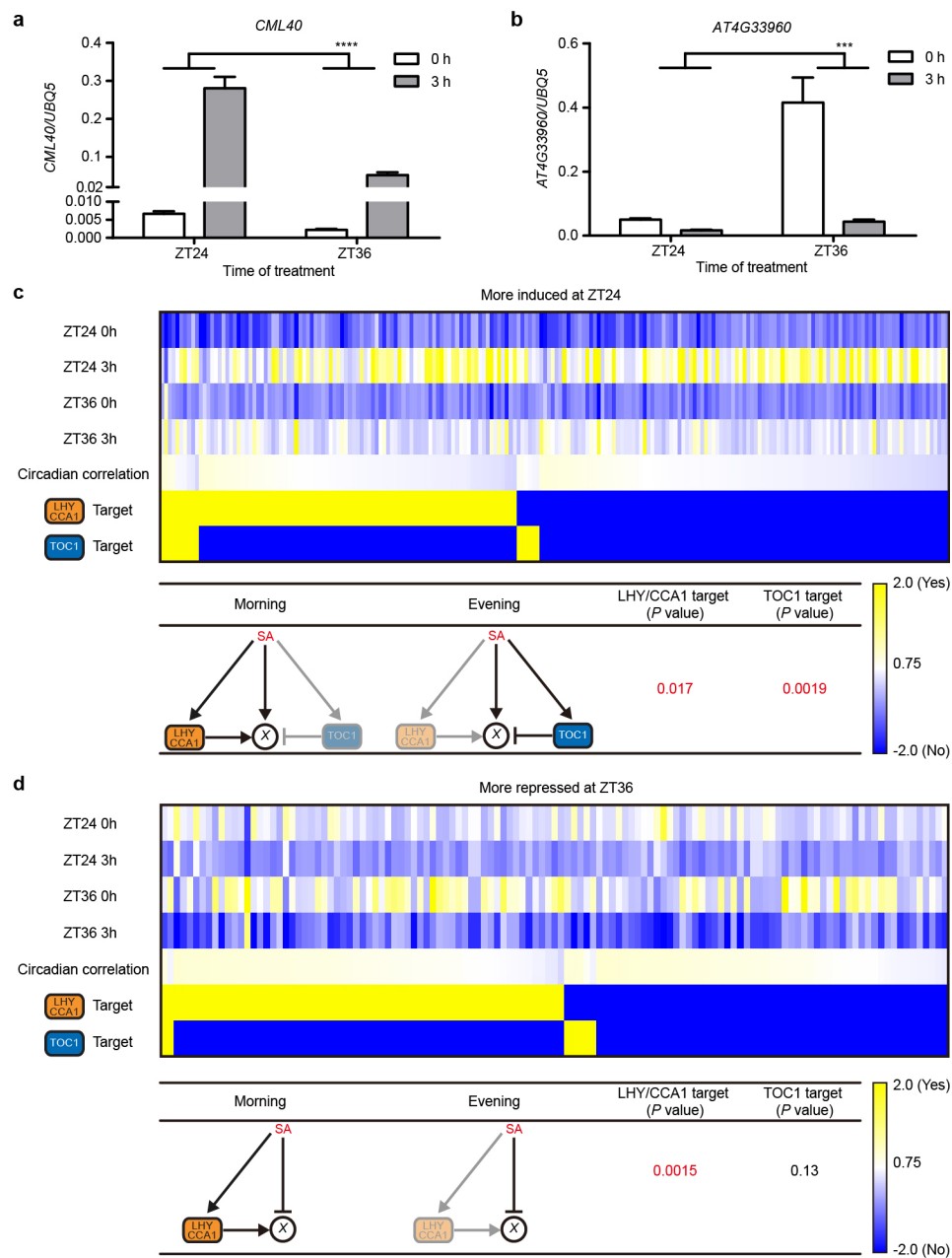
Extended Data Figure 5 | Redox perturbations affect the amplitude and average expression of *TOC1p:LUC* in an *NPR1*-dependent manner. **a**, *TOC1p:LUC* activity rhythms in 3-week-old soil-grown WT and *trx-h3 trx-h5* (*trx-h3 h5*) (mean \pm s.e.m., $n = 6$ plants). LL, constant light. White bars represent subjective days and grey bars represent subjective nights. The bar graphs show the estimates of amplitude and average expression (mean \pm s.e.m.; t -test; **** $P < 0.0001$). **b**, *TOC1p:LUC* activity rhythms in 3-week-old

soil-grown WT and *npr1* plants treated with water (CK) or 3 mM GSHmee at subjective dusk (black arrow) (mean \pm s.e.m., $n = 8$ plants). The bar graphs represent the estimates of amplitude and average expression of *TOC1p:LUC*, respectively (mean \pm s.e.m.). The letters above the bars indicate statistically significant differences between groups at $P < 0.01$ (Tukey's multiple comparisons test). ** $P < 0.01$; **** $P < 0.0001$ (two-way ANOVA). These experiments were repeated three times with similar results.



Extended Data Figure 6 | Model prediction and validation. **a**, Comparison of best-fit solutions for the *TOC1*-only and the *TOC1*-and-*PRR7* coupling in *npr1*. LL, constant light. White bars represent subjective days and grey bars represent subjective nights. **b**, Addition of *PRR7* coupling improves the fitness and mostly rescues the short period phenotype of the *TOC1*-only model (mean \pm s.e.m.; $n = 715$, n is degree of freedom derived from nonlinear regression). **c–e**, The transcript levels of *CCA1* (**c**), *LHY* (**d**), and *ELF3* (**e**) in

WT plants after water (CK) or 1 mM SA treatment. **f–h**, The transcript levels of *CCA1* (**f**), *LHY* (**g**), and *ELF3* (**h**) in WT and *npr1* plants. The expression was normalized to *UBQ5* (**c–h**). The bar graphs show the estimates of amplitude and average expression level, respectively (**c–h**; mean \pm s.e.m.; $n = 3$ biological replicates; t -test; $*P < 0.05$; $***P < 0.001$; $****P < 0.0001$). **i, j**, Comparison of best-fit solutions for NPR1 activation of *TOC1*-only (**i**) and NPR1 activation of *TOC1* and *LHY/CCA1* (**j**) after SA treatment.

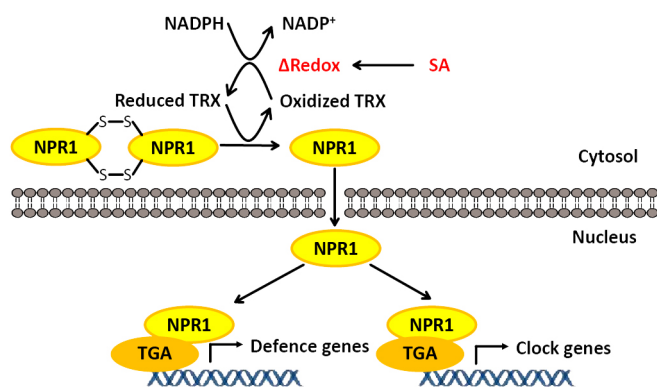


Extended Data Figure 7 | Validation and analysis of microarray data.

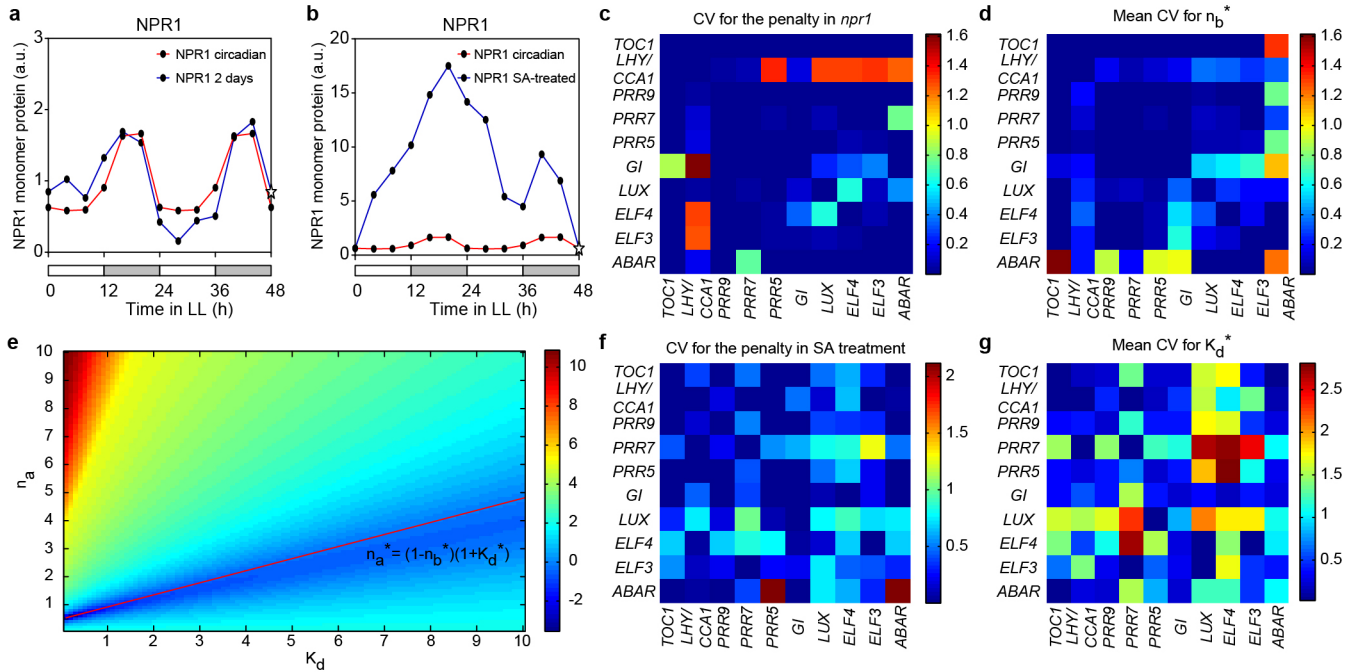
a, b, The transcript levels of *CML40* (**a**) and *AT4G33960* (**b**) in 3-week-old soil-grown plants 0 or 3 h after application of 1 mM SA either in the subjective morning (ZT24) or in the subjective evening (ZT36) normalized to *UBQ5* under constant light conditions. Data are mean \pm s.e.m. ($n = 3$ biological replicates; two-way ANOVA; *** $P < 0.001$; **** $P < 0.0001$).

c, d, Enrichment of *cis*-elements affecting time-of-day-specific sensitivity to induction. Promoter analysis of genes that were more induced by SA when treated at ZT24 (**c**) or more repressed by SA when treated at ZT36 (**d**).

The heat maps show the average expression levels based on the microarray. Circadian correlation coefficients were extracted from Diurnal (http://diurnal.mocklerlab.org/diurnal_data_finders/new). Yellow represents a high value or a target of CCA1/LHY or TOC1. Blue represents a low value or not a target of CCA1/LHY or TOC1. X represents a gene that was more induced by SA when treated at ZT24 (**c**) or more repressed by SA when treated at ZT36 (**d**). Arrows represent activation. Blocked arrows represent repression. P values were determined on the basis of hypergeometric distribution.



Extended Data Figure 8 | NPR1 senses and transduces redox signals to trigger transcriptional reprogramming. SA-triggered redox changes induce the oligomer-to-monomer switch of NPR1. The monomer then enters the nucleus and upregulates both defence genes and clock genes through interaction with TGA transcription factors.



Extended Data Figure 9 | Technical details for model fitting. **a**, Normalized NPR1 monomer abundance in mock-treated samples. The blue line presents the mean values from Fig. 2b, where the value at 48 h (marked with an open star) was inferred to be the same as that at 0 h. The red line represents the smoothed values used for modelling by averaging over 2 days to create a 1-day trace, which was then repeated over 2 days. The smoothed data were normalized, such that the time average of NPR1 was equal to 1. LL, constant light. White bars represent subjective days and grey bars represent subjective nights. **b**, SA-treated NPR1 monomer abundance. NPR1 monomer abundance after SA treatment from Extended Data Fig. 4b was normalized so that 0 h has the same value as the corresponding mock-treated NPR1 monomer level. On the basis of the assumption that the SA induction lasted for

2 days, the value of the last time point was inferred to be equal to the basal level (marked with an open star). **c**, Coefficient of variation (CV) of least-squares residual Σ for 15 different, random initial parameters for the model fitting of *npr1* data. **d**, Coefficient of variation of n_b^* for 15 different, random initial parameters for the model fitting of *npr1* data. **e**, Optimal n_a^*, K_d^* exhibit a linear relationship. $\log(\Sigma)$ was plotted as a function of n_a and K_d for mock-treated TOC1-only coupling (no query pairs). A 'low', linear Σ region is evident and is described by a simple analytical linear relationship, $n_b^* = 0.5689 \text{ h}^{-1}$. **f**, Coefficient of variation of Σ for 15 different, random initial parameters for the model fitting of SA-treated data. **g**, Coefficient of variation of K_d^* for 15 different, random initial parameters for the model fitting of SA-treated data.

Extended Data Table 1 | Primer sequences

Purpose	Primer name	Sequences
Transgenic plants	TOC1p_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGAGATCGCTCGGCTCAACAA
	TOC1p_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCAATTGTTTTGTTTGTCAATC
	TOC1p_Mut1	ATATTTTCTCCAAGAGTCCGTGGCCTTTTCTC
	TOC1p_Mut2	TTTTTATTGTCCACGGACTCTCCTTGGCCTAA
	CAT3p_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTACCCAAAGCTTCTGGCATTGTTTTGACTTTTGTGCG
	CAT3p_R	GGGGACCACTTTGTACAAGAAAGCTGGGTAGGTGATGATAGAAGTTGATGATCCCCCAATAGGCTT
	CAT2p_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTACAAGTAATCGATCATCCTTAAGTTTGGT
	CAT2p_R	GGGGACCACTTTGTACAAGAAAGCTGGGTAGGTTTGTGAGAAGAGAGCTTGGAGAGA
Y1H	TOC1p_P4P1R_F	GGGGACAACCTTTGTATAGAAAAGTTGGAGATCGCTCGGCTCAACAA
	TOC1p_P4P1R_R	GGGGACTGCTTTTTTGTACAAACTTGATTGTTTTGTTTGTCAATC
	TOC1p_Mut1	ATATTTTCTCCAAGAGTCCGTGGCCTTTTCTC
	TOC1p_Mut2	TTTTTATTGTCCACGGACTCTCCTTGGCCTAA
	TGA1_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGAATTCGACATCGACACAT
	TGA1_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCCGTTGGTTCACGATGTCGAGT
	TGA2_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCTGATACCAGTCCGAGA
	TGA2_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCTCTGGGTCGAGCAAGCCA
	TGA3_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAGATGATGAGCTCTTCT
	TGA3_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCAAGTGTGTTCTCGTGGACGAGC
	TGA4_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGAATACAACCTCGACACAT
	TGA4_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCCGTTGGTTCACGTTGCCTAGC
	TGA5_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGGAGATACTAGTCCAAGA
	TGA5_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCTCTTGGTCTGGCAAGCCA
	TGA6_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCTGATACCAGTTCAAGG
	TGA6_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCTCTTGGCCGGGCAAGCCA
	TGA7_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGATGAGTTCTTCTTCTCCA
	TGA7_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCAAGTTGGTCTTGTGGACGAGC
qPCR	TOC1_qP_F	AATAGTAATCCAGCGCAATTTTCTTC
	TOC1_qP_R	CTTCAATCTACTTTTCTTCGGTGCT
	LHY_qP_F	CGCTGCTTCGGTCTGGCCTT
	LHY_qP_R	TGTAGCAGCGGCAATGGCAGT
	PRR7_qP_F	CAGTCCACGAGCGGTATCTC
	PRR7_qP_R	CCAGGGCCAGATCACAGTTT
	CCA1_qP_F	TGACCGGTCTCGTGTGGCT
	CCA1_qP_R	ACTGCGGCGTGCAATTGGACT
	ELF3_qP_F	TGGCAAACTCGTCTGAAGGA
	ELF3_qP_R	GCCAAGTGAGATTACAGTCCAT
	PR1_qP_F	CTCATACACTCTGGTGGG
	PR1_qP_R	TTGGCACATCCGAGTC
	WRKY40_qP_F	ACAACGTCTTGAGGAAGCAAC
	WRKY40_qP_R	TCCGTTGAGCTACTCTCCGA
	CML40_qP_F	GAGCCACCAAGGCAAGGTAT
	CML40_qP_R	GTCCTCGAGCTCCAACGATT
	AT4G33960_qP_F	CGTCCAGATTGTTATGCGGC
	AT4G33960_qP_R	TGGAGAAGGGTAAGAAGCGG
ChIP	UBQ5_qP_F	GACGCTTCATCTCGTCC
	UBQ5_qP_R	GTAACGTAGGTGAGTCC
	TOC1 ChIP promoter_F	TGTCCACGTCATCTCCTTGG
	TOC1 ChIP promoter_R	AGCTTAATGGTGGGACTTGGG
	TOC1 ChIP coding region_F	GAGGCAAGACGAAGTCCCTG
	TOC1 ChIP coding region_R	GCTGCACCTAGCTTCAAGCA

Genetic modification of the diarrhoeal pathogen *Cryptosporidium parvum*

Sumiti Vinayak^{1*}, Mattie C. Pawlowicz^{1*}, Adam Sateriale^{1*}, Carrie F. Brooks¹, Caleb J. Studstill¹, Yael Bar-Peled¹, Michael J. Cipriano¹ & Boris Striepen^{1,2}

Recent studies into the global causes of severe diarrhoea in young children have identified the protozoan parasite *Cryptosporidium* as the second most important diarrhoeal pathogen after rotavirus^{1–3}. Diarrhoeal disease is estimated to be responsible for 10.5% of overall child mortality⁴. *Cryptosporidium* is also an opportunistic pathogen in the contexts of human immunodeficiency virus (HIV)-caused AIDS and organ transplantation^{5,6}. There is no vaccine and only a single approved drug that provides no benefit for those in gravest danger: malnourished children and immunocompromised patients^{7,8}. *Cryptosporidiosis* drug and vaccine development is limited by the poor tractability of the parasite, which includes a lack of systems for continuous culture, facile animal models, and molecular genetic tools^{3,9}. Here we describe an experimental framework to genetically modify this important human pathogen. We established and optimized transfection of *C. parvum* sporozoites in tissue culture. To isolate stable transgenics we developed a mouse model that delivers sporozoites directly into the intestine, a *Cryptosporidium* clustered regularly interspaced short palindromic repeat (CRISPR)/Cas9 system, and *in vivo* selection for aminoglycoside resistance. We derived reporter parasites suitable for *in vitro* and *in vivo* drug screening, and we evaluated the basis of drug susceptibility by gene knockout. We anticipate that the ability to genetically engineer this parasite will be transformative for

Cryptosporidium research. Genetic reporters will provide quantitative correlates for disease, cure and protection, and the role of parasite genes in these processes is now open to rigorous investigation.

Cryptosporidium infection occurs through faecal oral transmission of the environmentally resilient oocyst. The oocyst shelters four sporozoites that emerge in the small intestine and invade the epithelium. Although there is no tissue culture system for continuous passage, *C. parvum* development can be observed for 2–3 days by infecting human ileocaecal adenocarcinoma cells (HCT-8)¹⁰. To achieve transfection, sporozoites were excysted from oocysts purified from the faeces of experimentally infected calves using a protocol that mimics stomach and intestinal passage¹¹, and then electroporated before infection of HCT-8 cells (Fig. 1a). The transfection plasmids used here flanked a variety of reporter genes with candidate *C. parvum* 5' and 3' regulatory sequences derived from highly expressed housekeeping genes. We observed significant reporter activity 48 h after transfection using plasmids carrying nanoluciferase (Nluc; Fig. 1b), a small ATP-independent enzyme from deep sea shrimp¹², but not firefly luciferase or fluorescent proteins. Nluc luminescence correlated with the number of parasites and the amount of DNA used for transfection. Luminescence was also shown to require the presence of parasite-specific promoter elements and the introduction of DNA into parasites and not host cells (Fig. 1).

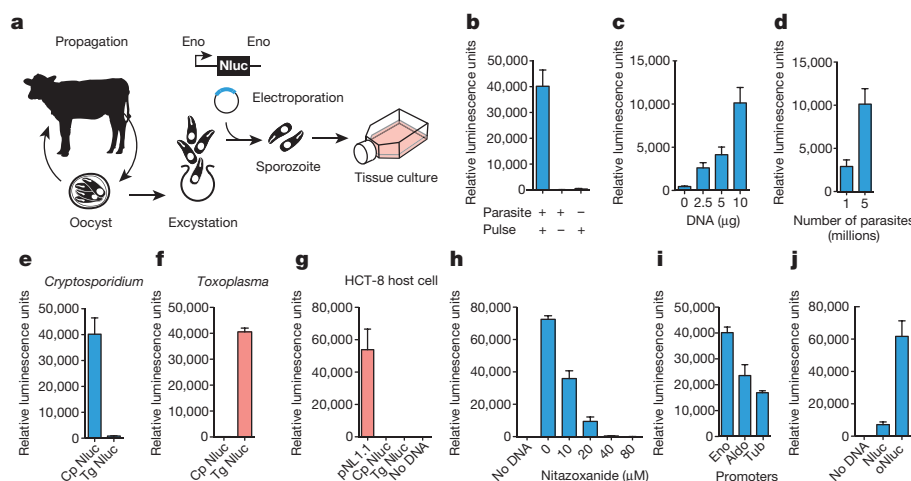


Figure 1 | Transfection of *C. parvum*. **a**, Schematic overview. *C. parvum* sporozoites were prepared from oocysts purified from infected calves and electroporated in the presence of plasmid DNA before infection of HCT-8 cells (Eno, flanking sequence from the *C. parvum* enolase gene). **b–j**, Luminescence measurements (the means of three technical replicates, standard deviation (s.d.) shown as error bars) of *C. parvum* (**b–e**, **h–j**, blue), *T. gondii* (**f**), or human HCT-8 cells (**g**) transfected with Nluc expression plasmids. **b–d**, *C. parvum* transfection requires electroporation (**b**) of DNA (**c**) into parasites (**d**). **e**, **f**, **h**, Transfection also requires plasmids to carry parasite-specific promoter sequences (**e**, **f**; testing *C. parvum* (Cp) and *T. gondii* (Tg) promoters in both

parasites), and is susceptible to the *Cryptosporidium* drug nitazoxanide (**h**). **g**, Lipofection of HCT-8 cells with the original Nluc plasmid pNL1.1 (Promega), but not derived parasite vectors, results in luciferase activity in the host alone. Choice of promoter (**i**; enolase (Eno), aldolase (Aldo), α -tubulin 5' regions (Tub) (the 3' untranslated region (UTR) was uniformly from the enolase gene)) or codon composition (**j**; Nluc optimized to 35% GC (oNluc)) influences expression level in *C. parvum*. Note automatic gain adjustment of luminescence measurements; units are not comparable between panels. Independent biological experiments were repeated three times, and representative data are shown.

¹Center for Tropical and Emerging Global Diseases, University of Georgia, Paul D. Coverdell Center, 500 D.W. Brooks Drive, Athens, Georgia 30602, USA. ²Department of Cellular Biology, University of Georgia, Paul D. Coverdell Center, 500 D.W. Brooks Drive, Athens, Georgia 30602, USA.

*These authors contributed equally to this work.

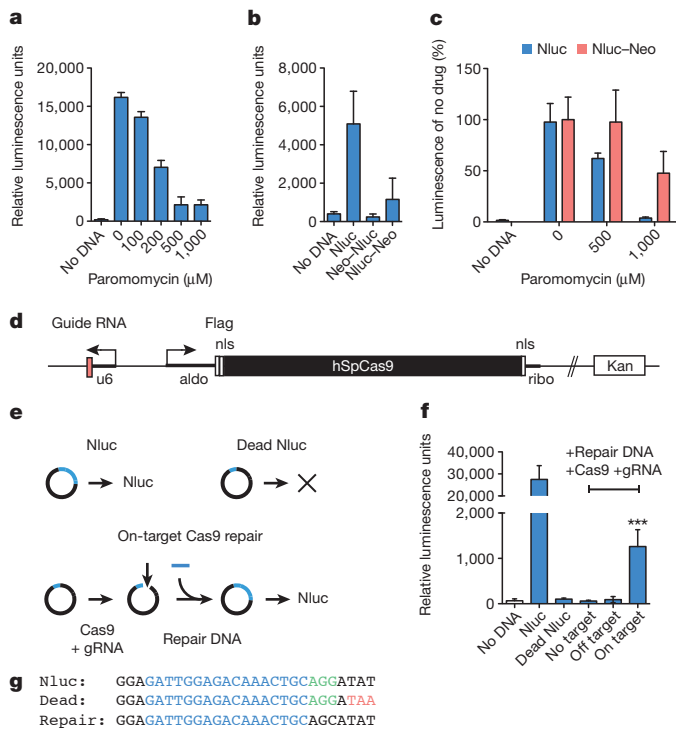


Figure 2 | Luciferase assays for *C. parvum* drug resistance and CRISPR/Cas9 activity. **a**, HCT-8 cells were infected with Nluc-transfected sporozoites and grown for 2 days in the presence of paromomycin. **b**, Translational fusions were constructed placing Neo at the amino or carboxy terminus of Nluc. Nluc-Neo shows luciferase activity, albeit at a reduced level when compared to Nluc alone. **c**, *C. parvum* transfected with Nluc (blue) or Nluc-Neo (red) were grown in different concentrations of paromomycin. Luciferase activity for each plasmid was normalized to its drug-free level. **d**, CRISPR/Cas9 plasmid for *C. parvum*. Flag, epitope tag; nls, nuclear localization signal; ribo, ribosomal protein L13A 3' UTR; u6, newly annotated promoter CM000433:553110–553472. **e**, **g**, Outline (**e**) and sequences (**g**) for Nluc repair assay. Guide RNA target, blue; protospacer adjacent motif, green; mutagenized codon 18, red. **f**, Sporozoites were transfected with Nluc or a codon 18 termination mutant (Dead Nluc); note ablation of signal. In addition to the Dead Nluc plasmid, some parasites also received a 125 bp double-stranded repair DNA fragment, and the Cas9 plasmid with the indicated guide RNAs (gRNAs; no target, empty gRNA cassette; off target, GFP gRNA; on target, Nluc gRNA). Statistical analysis compares Dead Nluc alone with Dead Nluc and Cas9 and specific gRNA. Note significant Cas9-mediated restoration of luciferase activity ($***P = 0.0006$, unpaired *t*-test). $n = 3$ technical replicates for **a–c**, and controls from **f**; $n = 6$ technical replicates for on-target samples in **f**. Error bars are s.d. and all experiments depicted here were repeated three times and representative data are shown.

Furthermore, reporter signal was ablated by the anti-parasitic drug nitazoxanide. Transient transfection of *C. parvum* is inefficient ($<10,000$ fold when compared to the related apicomplexan *Toxoplasma gondii* in parallel experiments) and requires a highly sensitive reporter such as Nluc to be noticeable.

In an effort to enhance efficiency we evaluated different electroporation devices, electrical wave programs and buffer compositions (Extended Data Fig. 1); this produced tenfold enhancement. We tested flanking sequences from different *C. parvum* genes and identified the enolase promoter to be strongest. The *C. parvum* genome is AT rich and shows strong codon bias¹³. We also noted a preference for A over T within the first 20 codons and thus explored codon optimization and found sixfold enhancement (Fig. 1j).

To enable enrichment of transgenic parasites, we next explored the selection of drug resistance. The aminoglycoside antibiotic paromomycin does not cure cryptosporidiosis in people, but is effective in tissue culture (Fig. 2a) and in immunocompromised mice¹⁴. Work in other protist models has shown aminoglycoside phosphotransferases to

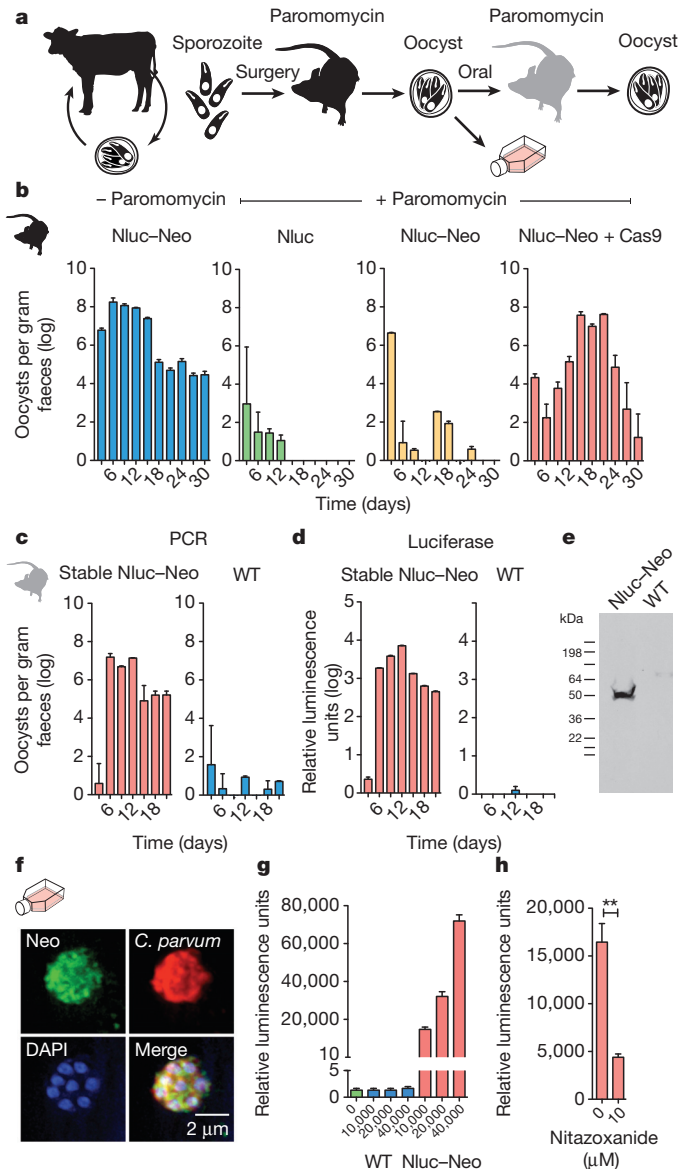


Figure 3 | Mouse model for selection of stable *C. parvum* transgenics.

a, Outline of the selection strategy. Transfected sporozoites were injected into the small intestine by surgery (Extended Data Fig. 2) and mice were treated with paromomycin. Oocysts were purified from the faeces and used to infect cultures or mice by oral gavage. **b**, Quantitative PCR of *C. parvum* DNA isolated from faeces of mice infected with transfected sporozoites (four mice per group) and treated as indicated. Emergence of paromomycin resistance required the Nluc-Neo and Cas9 plasmids. **c**, **d**, Upon reinfection, parasites show strong drug resistance (**c**) and luciferase activity (**d**). In repeat experiments we noted that luciferase is detectable as early as 6 days after transfection in the faeces of the first infected mouse (Extended Data Fig. 4). **e**, Protein extracts from oocysts were analysed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and western blot using an antibody against Neo (rabbit anti-neomycin phosphotransferase II; EMD Millipore). Predicted molecular mass of the Nluc-Neo fusion protein is 48.3 kDa. **f**, Immunofluorescence staining using anti-Neo (mouse anti-Neo; Alpha Diagnostic International) and *C. parvum* (tryptophan synthase B) antibodies. Note multiple nuclei in 4',6-diamidino-2-phenylindole (DAPI) stain typical for *C. parvum* meronts. No anti-Neo staining was observed in wild-type parasites. **g**, Luciferase assays for HCT-8 cultures infected with wild-type (WT; blue) and transgenic (Nluc-Neo; red) parasites. The y-axis is split to show level of luminescence background. $n = 3$ technical replicates, error bars are s.d., the experiment was done twice. **h**, Ninety-six-well luciferase drug assay using 1,000 oocysts per well. Note significant growth inhibition on treatment with 10 μ M nitazoxanide ($**P = 0.0036$, unpaired *t*-test). $n = 3$ technical replicates, error bars are s.d., the experiment was repeated two times and representative data are shown.

confer resistance to paromomycin^{15,16}. Appreciation of *C. parvum* drug resistance in culture is complicated by the lack of continuous growth. We thus constructed translational fusions between the Nluc reporter and the neomycin resistance marker (Neo)¹⁵ to focus our observation on the small subset of transfected parasites. Luciferase activity in parasites expressing Nluc-Neo showed reduced susceptibility to paromomycin treatment compared to Nluc alone (Fig. 2c), and thus we concluded that Nluc-Neo confers drug resistance in this transient assay.

Our genome searches indicated that *Cryptosporidium* species lack non-homologous end joining DNA repair. This suggested transgene integration to be rare and to require homologous recombination^{17,18}. Such recombination can be enhanced by long flanking regions and/or double-strand breaks introduced by restriction enzymes, transcription activator-like effector nucleases (TALENs) or CRISPR/Cas9 (refs 18, 19). To build a *C. parvum* CRISPR/Cas9 system, we constructed a plasmid in which the *C. parvum* U6 RNA promoter drives a guide RNA cassette²⁰ and the *Streptococcus pyogenes* Cas9 gene²¹ is flanked by parasite regulatory sequences (Fig. 2d). To test this system, we conducted a Cas9-dependent DNA repair experiment (Fig. 2e–g). We introduced a stop codon into the Nluc reporter that ablated luciferase activity (Dead Nluc). We then targeted the dead gene with a guide RNA, and provided a short double-stranded template for repair that restores read-through translation and renders the repaired gene resistant to further Cas9 cutting. When *C. parvum* sporozoites are co-transfected with a specific guide, luciferase activity is restored ($P = 0.0006$, unpaired *t*-test). No change is observed with no or off-target guides.

Interferon- γ knockout mice are susceptible to *C. parvum* infection through oral inoculation of oocysts²². However, infection with free sporozoites is less effective²³, probably due to stomach passage. We developed a surgical protocol to inject transfected sporozoites directly into the small intestine to maximize infection (Extended Data Fig. 2). When mice were killed 24 h after infection, luciferase activity was observed in scrapings of the intestinal epithelium. We also established an effective treatment protocol using paromomycin supplementation of the drinking water (Extended Data Fig. 3).

Next, we infected mice by surgery with transfected sporozoites and treated them with paromomycin as indicated (Fig. 3 and Extended Data Fig. 4; four mice per group). Faeces were collected every 3 days and oocyst shedding was measured by quantitative polymerase chain reac-

tion (PCR) targeting the *C. parvum* 18S ribosomal RNA locus. Mice infected with parasites transfected with the Nluc-Neo plasmid that did not receive drug shed high numbers of oocysts and remained infected for the 30 days observed (Fig. 3b, blue). Those infected with parasites that received the Nluc plasmid (lacking the Neo gene; Fig. 3b, green) were rapidly cured by drug treatment. Those transfected with Nluc-Neo alone and drug treated were also cured (infection may persist slightly longer). In contrast, infection with parasites carrying the Nluc-Neo plasmid and the Cas9 plasmid (Fig. 3b, red; Cas9 target detailed later) rapidly rebounded to levels similar to untreated mice. Oocysts emerging from selection were purified from faeces and used to infect mice that were again treated with paromomycin; wild-type oocysts were used in parallel (100,000 oocysts per mouse by gavage). While paromomycin treatment cured infection with wild-type parasites, transgenic parasites showed immediate robust drug resistance (Fig. 3c). When these oocysts were probed by western blot with anti-Neo antibody, we detected a band consistent with an Nluc-Neo fusion protein.

Purified oocysts were also used to infect cell cultures, and processed for immunofluorescence after 2 days. Transgenic but not wild-type intracellular parasite stages showed fluorescence when probed with antibodies specific for either Neo or Nluc (Fig. 3f and data not shown). These cultures also displayed strong luciferase activity not observed in wild type. This activity exceeded that previously observed in transient transfection experiments by five orders of magnitude on a per-cell basis. We assessed whether these organisms could be suitable for drug-screening assays by infecting 96-well plates with 1,000 oocysts per well and measured luciferase after 48 h. Infected wells were clearly distinguishable from uninfected wells ($z' > 0.6$; $n = 20$). Similarly, wells treated with nitazoxanide showed significant growth inhibition ($P = 0.0036$, unpaired *t*-test). Luciferase also provided a convenient way to assess the infection state of animals. We sampled 10 mg of faeces from mice diagnosed in parallel by PCR and found this assay to be sensitive, specific and faster than PCR (Fig. 3d). We note that Nluc expression remains stable when parasites are propagated in mice in the absence of paromomycin (Extended Data Fig. 5).

Cryptosporidium is remarkably resistant to antifolates, a mainstay of treatment against other apicomplexans, and this resistance has been attributed to differences in the target enzyme dihydrofolate reductase-thymidylate synthase (DHFR-TS)²⁴. However, *Cryptosporidium* is

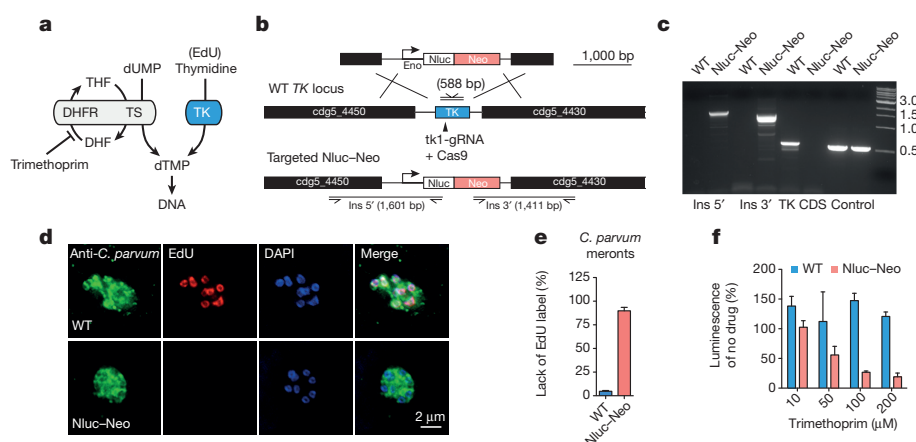


Figure 4 | Targeted deletion of *C. parvum* TK. **a**, Owing to a horizontal gene transfer, *C. parvum* has two pathways to synthesize dTMP: TK and DHFR-TS. DHF, dihydrofolic acid; THF, tetrahydrofolic acid; dUMP, uridine monophosphate. **b**, Map of the *C. parvum* TK locus, the targeting plasmid and the predicted modified locus. Primers and amplicon sizes of diagnostic PCR products are indicated (Ins, insertion). **c**, PCR analysis using genomic DNA from wild-type (WT) and transgenic parasites (Nluc-Neo, oocysts purified from faeces of infected mice shown in Fig. 3c; CDS, coding sequence). Primer sequences are provided in Supplementary Table 1. **e**, Quantification of EdU-labelling experiments (meronts with four or more nuclei were scored, two

biological repeats, $n = 105$ each sample, error bars are s.d.). **d**, Representative fluorescence micrographs are shown. Antibody to *C. parvum* tryptophan synthase B was used to identify parasites (green). **f**, Trimethoprim treatment of wild-type (blue) and Nluc-Neo transgenic (red) parasites. Wild-type parasites were measured in transient transfection assays with Nluc plasmid ($n = 3$, technical replicates, error bars are s.d.). The assay shown was conducted in the presence of 10 μM thymidine to avoid indirect host cell toxicity³⁰ (experiments without thymidine produced indistinguishable results). Experiments were repeated three times and representative data are shown.

unique among apicomplexans in that it acquired a thymidine kinase (TK) by horizontal gene transfer from bacteria²⁵. We hypothesized that TK may also contribute to *Cryptosporidium* antifolate resistance by providing an alternative route to thymidine monophosphate (dTMP; Fig. 4a). For this reason, the TK locus was targeted for insertion, allowing us to test this hypothesis by gene disruption. We mapped the locus in stable transgenic parasites by PCR using primers that link the marker genes with genomic sequences beyond the flanking regions on the targeting construct. This mapping is consistent with insertion by homologous double crossover (Fig. 4b, c). Furthermore, the TK coding sequence is no longer detectable, indicating uniform loss of the gene in the selected population. We tested for DNA incorporation of the thymidine analogue 5-ethynyl-2'-deoxyuridine (EdU) using click chemistry and fluorescence microscopy²⁶. Wild-type parasites grown in the presence of EdU show fluorescent nuclei. This labelling is lost in the transgenic parasites (Fig. 4d, e), confirming loss of TK at the biochemical level. We next treated parasite infected cultures with the antifolate trimethoprim. We confirmed the previously observed resistance in wild-type parasites, but noted enhanced susceptibility in the mutants (Fig. 4f). We conclude that the *C. parvum* TK is a non-essential enzyme required for the activation of thymidine, and that its presence limits the efficacy of antifolate therapy in *Cryptosporidium*.

We show that major hurdles towards genetic analysis and manipulation for cryptosporidiosis can be overcome by maximizing the efficiency of each step of the process and by focusing on *in vivo* propagation and selection. There is an urgent need for new anti-parasitic drugs³. *Cryptosporidium* is not susceptible to drugs widely used against related pathogens, which reflects substantial differences in its metabolism and metabolite uptake²⁷. Luciferase reporter parasites enable phenotypic screening in culture and animals with sufficient sensitivity and specificity to warrant a comprehensive effort to discover novel compounds. Gene deletion now permits biological target validation. Genetic modification may also allow the construction of attenuated parasites as a potential oral vaccine. While infants and toddlers are highly susceptible to the disease, infection is rarely detected in older children^{1,3}. This is consistent with infection studies in people and animals suggesting the development of anti-parasitic and anti-disease immunity^{3,28,29}. A better understanding of the mechanisms underlying disease and protection will be required to design and produce such a vaccine.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 May; accepted 12 June 2015.

Published online 15 July 2015.

- Kotloff, K. L. *et al.* Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet* **382**, 209–222 (2013).
- Mondal, D. *et al.* Contribution of enteric infection, altered intestinal barrier function, and maternal malnutrition to infant malnutrition in Bangladesh. *Clin. Infect. Dis.* **54**, 185–192 (2012).
- Checkley, W. *et al.* A review of the global burden, novel diagnostics, therapeutics, and vaccine targets for cryptosporidium. *Lancet Infect. Dis.* **15**, 85–94 (2015).
- Liu, L. *et al.* Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet* **379**, 2151–2161 (2012).
- Raja, K. *et al.* Prevalence of cryptosporidiosis in renal transplant recipients presenting with acute diarrhea at a single center in Pakistan. *J. Nephrol.* **3**, 127–131 (2014).
- Hunter, P. R. & Nichols, G. Epidemiology and clinical features of *Cryptosporidium* infection in immunocompromised patients. *Clin. Microbiol. Rev.* **15**, 145–154 (2002).
- Amadi, B. *et al.* Effect of nitazoxanide on morbidity and mortality in Zambian children with cryptosporidiosis: a randomised controlled trial. *Lancet* **360**, 1375–1380 (2002).
- Amadi, B. *et al.* High dose prolonged treatment with nitazoxanide is not effective for cryptosporidiosis in HIV positive Zambian children: a randomised controlled trial. *BMC Infect. Dis.* **9**, 195 (2009).

- Striemen, B. Parasitic infections: time to tackle cryptosporidiosis. *Nature* **503**, 189–191 (2013).
- Upton, S. J., Tilley, M. & Brillhart, D. B. Comparative development of *Cryptosporidium parvum* (Apicomplexa) in 11 continuous host cell lines. *FEMS Microbiol. Lett.* **118**, 233–236 (1994).
- Gut, J. & Nelson, R. G. *Cryptosporidium parvum*: synchronized excystation *in vitro* and evaluation of sporozoite infectivity with a new lectin-based assay. *J. Eukaryot. Microbiol.* **46**, 56S–57S (1999).
- Hall, M. P. *et al.* Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chem. Biol.* **7**, 1848–1857 (2012).
- Abrahamsen, M. S. *et al.* Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* **304**, 441–445 (2004).
- Theodos, C. M., Griffiths, J. K., D'Onfro, J., Fairfield, A. & Tzipori, S. Efficacy of nitazoxanide against *Cryptosporidium parvum* in cell culture and in animal models. *Antimicrob. Agents Chemother.* **42**, 1959–1965 (1998).
- Mochizuki, K. High efficiency transformation of *Tetrahymena* using a codon-optimized neomycin resistance gene. *Gene* **425**, 79–83 (2008).
- Gueiros-Filho, F. J. & Beverley, S. M. On the introduction of genetically-modified *Leishmania* outside the laboratory. *Exp. Parasitol.* **78**, 425–428 (1994).
- Fox, B. A., Ristuccia, J. G., Giggley, J. P. & Bzik, D. J. Efficient gene replacements in *Toxoplasma gondii* strains deficient for nonhomologous end-joining. *Eukaryot. Cell* **8**, 520–529 (2009).
- Lee, A. H., Symington, L. S. & Fidock, D. A. DNA repair mechanisms and their biological roles in the malaria parasite *Plasmodium falciparum*. *Microbiol. Mol. Biol. Rev.* **78**, 469–486 (2014).
- Brooks, C. F. *et al.* The *Toxoplasma* apicoplast phosphate translocator links cytosolic and apicoplast metabolism and is essential for parasite survival. *Cell Host Microbe* **7**, 62–73 (2010).
- Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
- Sidik, S. M., Hackett, C. G., Tran, F., Westwood, N. J. & Lourido, S. Efficient genome engineering of *Toxoplasma gondii* using CRISPR/Cas9. *PLoS ONE* **9**, e100450 (2014).
- Griffiths, J. K., Theodos, C., Paris, M. & Tzipori, S. The gamma interferon gene knockout mouse: a highly sensitive model for evaluation of therapeutic agents against *Cryptosporidium parvum*. *J. Clin. Microbiol.* **36**, 2503–2508 (1998).
- Fayer, R., Nerad, T., Rall, W., Lindsay, D. S. & Blagburn, B. L. Studies on cryopreservation of *Cryptosporidium parvum*. *J. Parasitol.* **77**, 357–361 (1991).
- Liu, J., Bolstad, D. B., Bolstad, E. S. D., Wright, D. L. & Anderson, A. C. Towards new antifolates targeting eukaryotic opportunistic infections. *Eukaryot. Cell* **8**, 483–486 (2009).
- Striemen, B. *et al.* Gene transfer in the evolution of parasite nucleotide biosynthesis. *Proc. Natl Acad. Sci. USA* **101**, 3154–3159 (2004).
- Salic, A. & Mitchison, T. J. A chemical method for fast and sensitive detection of DNA synthesis *in vivo*. *Proc. Natl Acad. Sci. USA* **105**, 2415–2420 (2008).
- Striemen, B. in *Antimicrobial Drug Resistance* Vol. 1 (eds Mayers, D. L., Lerner, S. A., Quellet, M. & Sobel, J. D.) 605–621 (Springer, 2009).
- Sheoran, A., Wiffin, A., Widmer, G., Singh, P. & Tzipori, S. Infection with *Cryptosporidium hominis* provides incomplete protection of the host against *Cryptosporidium parvum*. *J. Infect. Dis.* **205**, 1019–1023 (2012).
- McDonald, V., Deer, R., Uni, S., Iseki, M. & Bancroft, G. J. Immune responses to *Cryptosporidium muris* and *Cryptosporidium parvum* in adult immunocompetent or immunocompromised (nude and SCID) mice. *Infect. Immun.* **60**, 3325–3331 (1992).
- Jiang, L., Lee, P. C., White, J. & Rathod, P. K. Potent and selective activity of a combination of thymidine and 1843U89, a folate-based thymidylate synthase inhibitor, against *Plasmodium falciparum*. *Antimicrob. Agents Chemother.* **44**, 1047–1050 (2000).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Sharling for initial contributions and L. Hedstrom, J. Mead, S. Vaishnav, L. Xiao and Y. Belkaid for discussion. This work was funded in part by the National Institutes of Health (NIH; R01AI112427) to B.S. and by a pilot grant from the Centers for Disease Control and the University of Georgia Research Foundation to B.S. and L. Xiao. M.J.C. was supported by training grant NIH T32AI060546 and B.S. is a Georgia Research Alliance Distinguished Investigator.

Author Contributions S.V. developed the transfection and luciferase assay; M.C.P. optimized transfection and developed the Cas9 system; S.V., M.C.P., A.S. and C.F.B. developed the mouse infection protocol and A.S. developed selection assays; C.F.B. developed surgery; C.J.S. and Y.B.-P. constructed some of the plasmids; and M.J.C. provided bioinformatics support. S.V., M.C.P., A.S. and C.F.B. conducted animal experiments and genotypic and phenotypic characterization. S.V., M.C.P., A.S., C.F.B. and B.S. conceived the study and B.S. wrote the manuscript with contributions from S.V., M.C.P. and A.S.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.S. (striemen@uga.edu).

METHODS

***C. parvum* reporter and drug resistance vectors.** *C. parvum* transfection vectors were derived from plasmid pH₃BG¹⁹ and modified to contain *C. parvum* promoter and 5' and 3' untranslated messenger RNA regions. We mined the genome and a variety of expression data sets collectively available through Crypto DB (<http://www.cryptodb.org>)³¹ to identify genes that are highly expressed across the life-cycle. Promoters and 5' UTRs of the enolase (cgd5_1960), α -tubulin (cgd4_2860), and aldolase (cgd1_3020) genes and 3' UTRs of enolase (51 bp), α -tubulin (97 bp) or ribosomal protein L13A (cgd5_970, UTR 211 bp) were amplified from genomic DNA by PCR (see Supplementary Table 1 for a list of primer sequences and restriction sites used). Nluc was amplified from pNL1.1 (Promega Corporation), firefly luciferase and different fluorescent protein genes were amplified from vectors used for *T. gondii*^{32–34}. The neomycin resistance gene was amplified from plasmid pNeo4 (ref. 15) (a gift from J. Gaertig, University of Georgia) and introduced 5' or 3' of Nluc in a plasmid with enolase regulatory sequences. To target the *TK* gene, regions flanking the gene were amplified and introduced into the Nluc–Neo vector (the promoter but not the 3' UTR was retained).

***C. parvum* CRISPR/Cas9 genome editing.** Human codon-optimized *Streptococcus pyogenes* Cas9 (hSpCas9) carrying a Flag tag and N- and C-terminal nuclear localization signals was amplified from pX330 (ref. 35) and introduced into the Aldolase–Nluc–ribo vector replacing the Nluc. A guide RNA cassette was synthesized containing the *C. parvum* U6 promoter identified by genome searches using known structural RNA sequences from *Plasmodium falciparum*³⁶, two inverted BbsI restriction sites to facilitate guide cloning, a *trans*-activating CRISPR RNA (tracrRNA) consensus sequence and a terminator (poly T) sequence, and was introduced into the Cas9 plasmid.

To test for CRISPR/Cas9-mediated repair *in vitro*, we modified the codon-optimized Nluc vector by introducing a premature stop codon (Y18Stop) adjacent to a guide target sequence at the beginning of the gene by site-directed mutagenesis (QuikChange II, Agilent Technologies). A 125 bp double-stranded (ds)DNA oligonucleotide was synthesized that restored Y18 and disrupted the PAM motif (G17A) of the guide RNA target, thus rendering it resistant to further Cas9 cuts.

Parasite excystation and transfection. Oocyst excystation was carried out as described¹¹ with some modification. Up to 10⁸ *C. parvum* Iowa strain oocysts (Sterling Parasitology Laboratory or Bunch Grass Farm) were suspended in 100 μ l of 1:4 aqueous dilution of 5.25% sodium hypochlorite and incubated on ice for 5 min. Oocysts were then washed three times with ice-cold PBS, suspended at 3.9 \times 10⁵ oocysts per ml of 0.2 mM sodium taurocholate (prepared in PBS) and incubated at 15 °C (10 min) and then at 37 °C (60–90 min). Emergence of sporozoites was monitored microscopically (typical efficiency 70–90%). Sporozoites were filtered through a 3 μ m polycarbonate filter to remove unexcysted oocysts, washed with ice-cold PBS, and counted.

Initially we used a BTX ECM 630 device for electroporation (Harvard Apparatus). Excysted sporozoites (10⁷) were suspended in complete cytomix buffer (120 mM KCl, 0.15 mM CaCl₂, 10 mM K₂HPO₄/KH₂PO₄, pH 7.6, 25 mM HEPES, pH 7.6, 2 mM EGTA, 5 mM MgCl₂, pH 7.6, supplemented with 2 mM ATP and 5 mM glutathione), mixed with plasmid DNA, and electroporated with a single 1,500 V pulse, resistance of 25 Ω , and a capacitance of 25 μ F. To enhance transfection efficiency, we switched to using the AMAXA Nucleofactor 4D device (Lonza Cologne GmbH). After excystation, 10⁷ sporozoites were suspended in 15 μ l Lonza SF Buffer and combined with 10–50 μ g DNA (prepared in Tris–EDTA, pH 8.0) at a final volume of 20 μ l. The parasite–DNA mix was added to small, strip cuvettes and electroporated using program EH100. Additional electroporation conditions were explored to arrive at this protocol and those are listed in Extended Data Fig. 1.

For *in vitro* transfection assays, human ileocaecal adenocarcinoma (HCT-8) cells (ATCC) were grown in RPMI-1640 with glutamine supplemented with 10% FBS, 1 mM sodium pyruvate, 50 U ml^{−1} penicillin, 50 μ g ml^{−1} streptomycin and amphotericin B in 24-, 48- or 96-well plates to 70% confluency. No effort was made to authenticate this cell line or test for mycoplasma. Prior to infection, media was replaced with DMEM with 2% FBS, 50 U ml^{−1} penicillin, 50 μ g ml^{−1} streptomycin and amphotericin B, and 0.2 mM L-glutamine. For *in vivo* experiments electroporated sporozoites were suspended in PBS and kept on ice until administered to the mice.

The *T. gondii* Nluc plasmid was constructed by inserting the Nluc sequence into vector pCTH₃ (ref. 32) and parasites were electroporated and used to infect human foreskin fibroblasts as described³⁷. HCT-8 cells were cultured in 24-well plates until confluent, transfected with 500 ng of DNA using Lipofectamine 2000 as described by the manufacturer (Life Technologies), and assayed for Nluc activity after 48 h.

Animal ethics statement. Animal experiments were approved by the Institutional Animal Care and Use Committee of the University of Georgia (animal use protocol no. A2012 03-028-Y3-A12).

Surgical delivery of transfected sporozoites into IFN- γ -deficient mice. In preliminary experiments we noted that antibiotic removal of bacterial flora enhances susceptibility of mice. Prior to infection mice were orally treated by gavage daily for a week before infection with an antibiotic cocktail (3 mg ampicillin, 3 mg streptomycin, 0.95 mg metronidazole, 3 mg neomycin and 1.5 mg vancomycin in distilled H₂O, per mouse/per day; all antibiotics purchased from Sigma). To deliver sporozoites directly to the small intestine, we developed a mouse survival surgery protocol for female C57BL/6 IFN- γ -deficient mice (B6.129S7-Ifng^{tm1Ts}/J, Jackson Laboratories) aged 6–8 weeks. The abdominal area of mice was shaved with clippers. Animals were placed in an isoflurane (3–5%) anaesthesia induction chamber and then moved to a nosecone (1–3% isoflurane as needed) on a sterile surgical field. A sterile drape was applied over a warming pad after sterilization of the area with 70% ethanol. Respiration and response to stimulation (toe pinch) were monitored during the procedure and the vaporizer adjusted as needed. Mucous membranes and footpads were monitored for colour to confirm adequate perfusion. Three betadine (Povidone-iodine) scrubs followed by a 70% ethanol wipe were applied to shaved skin before surgery. Ophthalmic ointment (Puralube, Dechra Veterinary Products) was applied to prevent drying of eyes. Skin was vertically incised midline of the abdominal region below the sternum with microsurgical scissors for approximately 1.5 cm followed by vertical incision of the peritoneum. Exposed jejunum/ileum was injected with 10⁷ transfected sporozoites suspended in 200 μ l PBS containing sterile food colouring dye as tracer. After injection, suturing was performed to close the peritoneum. Mice were administered 0.01–0.02 ml per gram body weight of warm lactated Ringer's solution subcutaneously after surgery. Meloxicam analgesic was also administered to the mice after surgery. At completion of the procedure, the eye ointment was wiped off and the vaporizer was turned off and the mice were allowed to breathe the oxygen supply gas until they began to wake. Mice were placed in a recovery area until ambulatory and exhibiting normal respiration and were watched for 2 h after surgery. Incision sites were monitored daily until fully healed (10–14 days). Twenty-four hours after surgical infection, water in mouse cages was replaced with distilled H₂O containing 16 mg ml^{−1} paromomycin, a concentration we determined to deliver a daily dose of 40 mg kg^{−1} paromomycin to each mouse (Extended Data Fig. 3). Mice were randomly assigned to groups before surgery. A sample size of four animals per treatment group was judged to be sufficiently large enough to draw appropriate conclusions. All mice survived surgery and were included in the results reported here. Investigators were not blinded to group allocation during the experiments.

Mouse faeces collection and storage. Faecal samples were collected from mice (typically four mice per cage) starting 3 days after infection every third day for up to a month. Mice were transferred to a fresh, sterile cage for 2–3 h, and faeces from the cage were collected, pooled, and stored at 4 °C.

Luciferase assay. For transient transfection experiments, electroporated sporozoites were added to 70% confluent HCT-8 culture and infection was allowed to proceed at 37 °C for 48 h. Media was removed from wells and 200 μ l of NanoGlo lysis buffer supplemented with NanoGlo substrate (1:50, Promega Corporation) was added to each well. Cells were scraped and the lysate was transferred to white 96-well plates and luminescence was measured using a Synergy H4 Hybrid Microplate Reader (BioTek Instruments). For drug assays with purified stable transgenic oocysts, the culture supernatant was collected after 48 h from 96-well plates. An equal amount of supernatant and NanoGlo lysis buffer with substrate was combined and luminescence was measured.

For luciferase measurement from mouse faecal samples, 20 mg of faeces was weighed into a 1.5-ml microcentrifuge tube and homogenized in 1 ml of lysis buffer (50 mM Tris–HCl, 10% glycerol, 1% Triton-X, 2 mM dithiothreitol (DTT), 2 mM EDTA) using 10–15 glass beads (3 mm) and a vortex mixer for 1 min, followed by clarification of lysate by brief centrifugation. One-hundred microlitres of lysate was mixed with an equal volume of NanoGlo Luciferase Buffer (prepared with 1:50 dilution of substrate) and luminescence was measured as described.

High-throughput imaging assay for parasite growth. For drug assays we used either luciferase activity or a 96-well infection and imaging protocol³⁸ using a BD Pathway instrument. Parasites and host cells were quantified using an ImageJ macro adapted from ref. 39. The ratio of parasites to host nuclei was determined for each sample image and normalized to untreated controls.

For oocyst quantification by high-throughput microscopy, we weighed collected mouse faeces and diluted in PBS (5 μ l mg^{−1}). Samples were incubated at 95 °C for 10 min, vortexing every 2 min at high speed. Large debris was allowed to settle for 10 min, then 10 μ l of the suspension were mixed with 990 μ l PBS and 1 μ l of fluorescein isothiocyanate (FITC)-conjugated goat polyclonal anti-*Cryptosporidium* antibody (GeneTex). After 1 h at room temperature, the sample was centrifuged at 2,000g for 15 min. The pellet was suspended in 200 μ l PBS and transferred to a 96-well plate for microscopy. Plates were imaged using BD Pathway and oocysts were counted

using an ImageJ macro. Using a standard curve (uninfected mouse faeces spiked with known amounts of oocysts), oocyst counts were converted to oocysts per grams faeces.

Quantification of oocyst shedding using qRT-PCR. DNA was extracted from 100 mg faeces using ZR Faecal DNA MiniPrep Kit (Zymo Research Corporation) following the manufacturer's protocol with slight modification. While in lysis buffer, the sample was freeze-thawed in liquid nitrogen five times before the first centrifugation step. Each sample was eluted in 50 µl water, 1 µl of eluate was used for qRT-PCR along with 10 µM primers targeting *Cryptosporidium* 18S rRNA⁴⁰ and SYBR Master Mix (Life Technologies) for detection. Each qRT-PCR reaction was normalized using an eight-point standard curve (faecal DNA purified from uninfected mouse faeces spiked with known amounts of oocysts) for each set of samples.

Oocyst purification from mouse faeces. Oocysts were purified from faeces using sucrose suspension followed by a caesium chloride centrifugation⁴¹. Mouse faeces were suspended in tap water, passed through a 850-µm mesh filter, followed by 250-µm mesh. This filtered suspension was mixed 1:1 with aqueous sucrose solution (specific gravity 1.33), and centrifuged at 1,000g for 5 min. Oocysts were collected from the supernatant and suspended in 0.85% saline solution. 0.5 ml of this preparation was overlaid onto 0.8 ml of 1.15 specific gravity CsCl, and centrifuged for 3 min at 16,000g. Oocysts were collected from the top ml of the sample, washed in 0.85% saline, counted with disposable counting chamber (Kova International) and suspended in 2.5% potassium dichromate for storage at 4 °C.

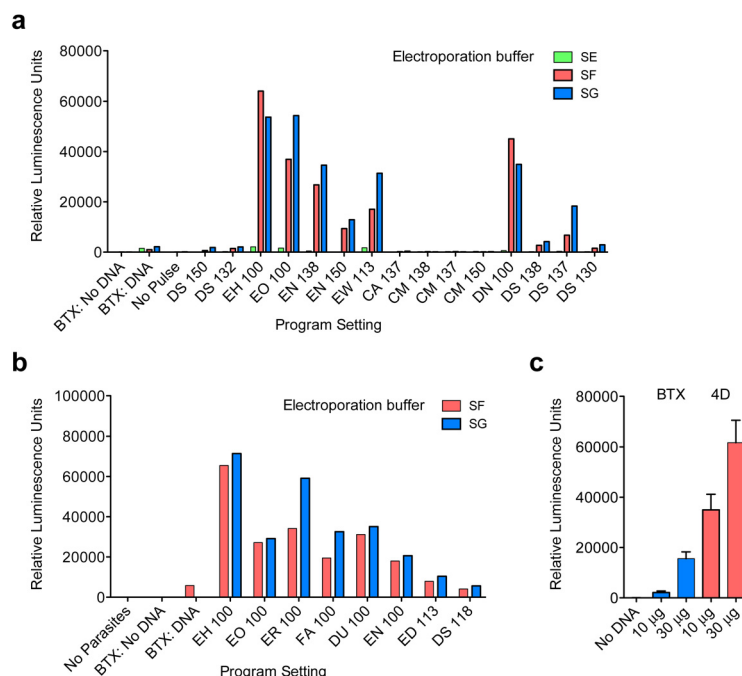
Western blotting. For western blot analysis, oocysts from wild-type and transgenic Nluc-Neo parasites were excysted as described earlier and sporozoites were lysed in SDS sample buffer. Protein extract from 10⁷ sporozoites was loaded per lane and subjected to electrophoresis on a precast Any kD Mini-PROTEAN TGX gel (Bio-Rad) followed by transfer to 0.2-µm nitrocellulose membrane (Bio-Rad). Blots were blocked and probed with an anti-neomycin phosphotransferase II antibody (EMD Millipore) at 1:1,000 dilution and goat anti-rabbit IgG (H + L)-HRP conjugate (Bio-Rad) at 1:20,000 dilution followed by detection with ECL Western Blotting Substrate (Thermo Pierce) and exposure to film. Equal loading of blots was controlled by stripping and reprobing with an antibody to α -tubulin.

EdU labelling and immunofluorescence microscopy. EdU labelling was performed using the Click-iT EdU Alexa Fluor 594 Imaging Kit following the manufacturer's instructions (Life Technologies). Purified stable transgenic oocysts expressing the luciferase or wild-type oocysts were inoculated into 24-well plates containing coverslips confluent with HCT-8 cells. After 24 h, EdU was added to the media at 10 µM and left for 18 h before fixation. For immunofluorescence, primary antibodies used were mouse monoclonal anti-human neomycin phosphotransferase II (NPII) (Alpha Diagnostic International), rabbit polyclonal

anti-Nluc antibody (Promega Corporation), and polyclonal rabbit anti-*C. parvum* tryptophan synthase B (TrpB; B.S., unpublished observations) at 1:1,000, secondary antibodies were anti-mouse or anti-rabbit conjugated to Alexa488 or Alexa546 (Molecular Probes, Life Technologies) at a dilution of 1:1,000. DNA was visualized with DAPI (2 mg ml⁻¹). Images were collected on an Applied Precision Delta Vision inverted epifluorescence microscope at the UGA Biomedical Microscopy Core, deconvolved and adjusted for contrast using SoftWoRx software.

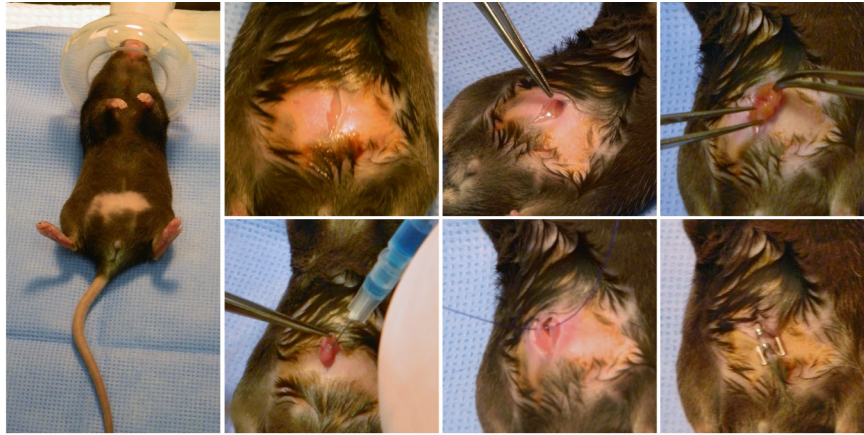
Statistical methods. All bar graphs depict the mean with standard deviations shown as error bars. Unless indicated otherwise, graphed data represent three technical replicates; each experiment was repeated at least twice and representative data are shown. No statistical tests were used to predetermine sample size. Unpaired *t*-tests were used appropriately to determine statistical significance and a *P* value <0.05 was considered significant. Assumptions for statistical tests were confirmed or corrected as described. No animals were excluded from experimental measurements.

31. Harb, O. S. & Roos, D. S. The Eukaryotic Pathogen Databases: a functional genomic resource integrating data from human and veterinary parasites. *Methods Mol. Biol.* **1201**, 1–18 (2015).
32. van Dooren, G. G., Tomova, C., Agrawal, S., Humbel, B. M. & Striepen, B. *Toxoplasma gondii* Tic20 is essential for apicoplast protein import. *Proc. Natl Acad. Sci. USA* **105**, 13574–13579 (2008).
33. Gubbels, M. J., Li, C. & Striepen, B. High-throughput growth assay for *Toxoplasma gondii* using yellow fluorescent protein. *Antimicrob. Agents Chemother.* **47**, 309–316 (2003).
34. Saeij, J. P., Boyle, J. P., Grigg, M. E., Arrizabalaga, G. & Boothroyd, J. C. Bioluminescence imaging of *Toxoplasma gondii* infection in living mice reveals dramatic differences between strains. *Infect. Immun.* **73**, 695–702 (2005).
35. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
36. Chakrabarti, K. *et al.* Structural RNAs of known and unknown function identified in malaria parasites by comparative genomics and RNA analysis. *RNA* **13**, 1923–1939 (2007).
37. Striepen, B. & Soldati, D. in *Toxoplasma gondii: The Model Apicomplexan — Perspective and Methods* (eds Weiss, L. M. & Kim, K.) 391–415 (Elsevier, 2007).
38. Sharling, L. *et al.* A screening pipeline for antiparasitic agents targeting cryptosporidium inosine monophosphate dehydrogenase. *PLoS Negl. Trop. Dis.* **4**, e794 (2010).
39. Besoff, K., Sateriale, A., Lee, K. K. & Huston, C. D. Drug repurposing screen reveals FDA-approved inhibitors of human HMG-CoA reductase and isoprenoid synthesis that block *Cryptosporidium parvum* growth. *Antimicrob. Agents Chemother.* **57**, 1804–1814 (2013).
40. Mary, C. *et al.* Multicentric evaluation of a new real-time PCR assay for quantification of *Cryptosporidium* spp. and identification of *Cryptosporidium parvum* and *Cryptosporidium hominis*. *J. Clin. Microbiol.* **51**, 2556–2563 (2013).
41. Upton, S. J. in *Cryptosporidium and Cryptosporidiosis* (ed. Fayer, R.) 181–207 (CRC, 1997).



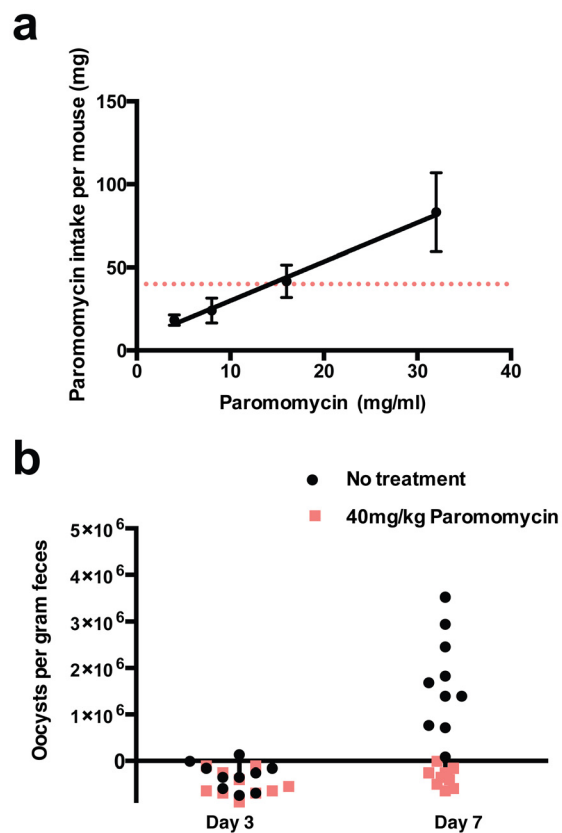
Extended Data Figure 1 | Optimization of sporozoite transfection. a, Ten-million sporozoites prepared in either cytomix (BTX) or Lonza Buffers SE, SF or SG (4D Nucleofection) were combined with 10 μ g DNA (Eno_Nluc-GS-Nluc_Eno). Samples were electroporated using previously determined settings for BTX (1,500 V, 25 Ω , 25 μ F) or various program settings for 4D Nucleofection as indicated. Parasites were added to cultures of HCT-8 cells and luciferase activity was read after 48 h. Bars represent average of two technical replicates. **b,** Transfection was further optimized by comparing the best preliminary settings (buffers SF and SG; programs EH 100 and EO 100) with additional pulse programs as indicated. Transfection was carried out as in

a. Bars represent average of two technical replicates. **c,** Electroporation systems (BTX and 4D Nucleofection) were compared using the same number of *C. parvum* sporozoites and quantities of DNA using buffers and conditions optimized in **a** and **b**. Bars represent average of three technical replicates. Note about tenfold enhancement of transient transfection using 4D Nucleofection. The impact of electroporation on stable transformation cannot be assessed in this setup and may be higher. Experiments in **a** and **b** were done once for the purpose of optimization, while **c** was repeated three times; a single representative experiment is shown.

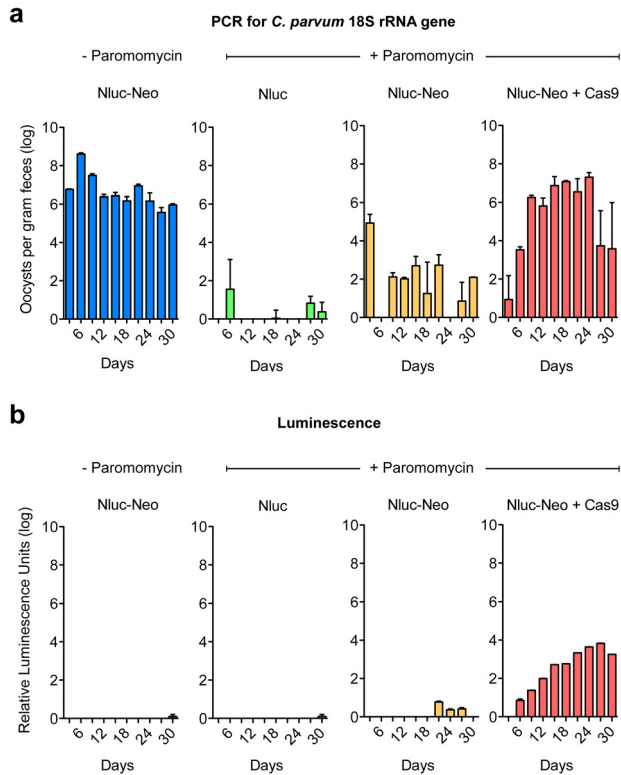


Extended Data Figure 2 | Direct surgical injection of transfecting *C. parvum* sporozoites into the small intestine. Mice are shaved and anaesthetized with isoflurane (3% initially, then maintained at 1.5% for the surgery). The abdominal skin is disinfected with Betadine and a small incision is made into the peritoneum. Forceps are used to grasp the small intestine and 100 μ l of

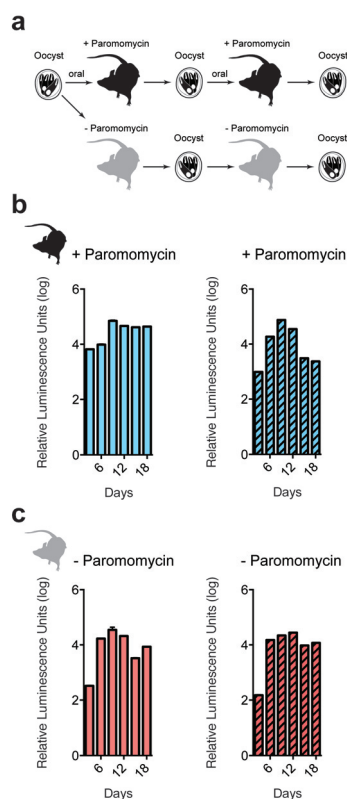
PBS containing 10^7 transfecting *C. parvum* sporozoites is injected into the lumen. The peritoneum and the abdominal skin are each sutured with 4-0 polydioxanone and mice are injected with meloxicam (1 mg kg^{-1}) subcutaneously. Each procedure takes around 15 min, and mice recover rapidly.



Extended Data Figure 3 | Optimization of paromomycin treatment of infected mice. **a**, Dosing of mice accounting for drug concentration, animal weight, and measured daily water consumption. At 16 mg ml^{-1} each mouse received 40 mg paromomycin daily (dotted line). **b**, This dose was found to be sufficient to decrease oocyst shedding in treated mice to background. By day 7 mice without paromomycin treatment shed large amounts of oocysts when compared to untreated mice. Treated mice showed no shedding above background. Oocysts were enumerated by high-throughput imaging assay. Five mice were analysed individually with two technical replicates.



Extended Data Figure 4 | Mouse model for selection of stable *C. parvum* transgenics. Repeat of the experiment described in Fig. 3b. **a**, Measurement of *C. parvum* infection using faecal PCR. **b**, Luminescence measurements. Note increasing luminescence from day 6 in parasites that received resistance and Cas9 plasmids. Mice were infected in groups of four per cage and pooled faeces was analysed for each cage (each measurement represents three technical replicates).



Extended Data Figure 5 | *C. parvum* maintains the stable transgene when passed serially in mice without paromomycin treatment. **a**, Mice were infected orally with 100,000 transgenic oocysts. **b, c**, Infected mice were then treated with paromomycin (**b**) or left untreated (**c**). Oocysts were purified from faecal collections by sucrose flotation and CsCl centrifugation, and used to infect a second cohort of mice. Again, each mouse received 100,000 transgenic oocysts and mice were treated or not. Faeces were tested for luminescence every 3 days. Each reading represents the pooled faecal sample from five mice with three technical replicates.

Engineered CRISPR–Cas9 nucleases with altered PAM specificities

Benjamin P. Kleinstiver^{1,2,3}, Michelle S. Prew^{1,2}, Shengdar Q. Tsai^{1,2,3}, Ved V. Topkar^{1,2}, Nhu T. Nguyen^{1,2}, Zongli Zheng^{1,3,4}, Andrew P. W. Gonzales^{5,6,7}, Zhuyun Li⁵, Randall T. Peterson^{5,6,7}, Jing–Ruey Joanna Yeh^{5,8}, Martin J. Aryee^{1,3,9} & J. Keith Joung^{1,2,3}

Although CRISPR–Cas9 nucleases are widely used for genome editing^{1,2}, the range of sequences that Cas9 can recognize is constrained by the need for a specific protospacer adjacent motif (PAM)^{3–6}. As a result, it can often be difficult to target double-stranded breaks (DSBs) with the precision that is necessary for various genome-editing applications. The ability to engineer Cas9 derivatives with purposefully altered PAM specificities would address this limitation. Here we show that the commonly used *Streptococcus pyogenes* Cas9 (SpCas9) can be modified to recognize alternative PAM sequences using structural information, bacterial selection-based directed evolution, and combinatorial design. These altered PAM specificity variants enable robust editing of endogenous gene sites in zebrafish and human cells not currently targetable by wild-type SpCas9, and their genome-wide specificities are comparable to wild-type SpCas9 as judged by GUIDE-seq analysis⁷. In addition, we identify and characterize another SpCas9 variant that exhibits improved specificity in human cells, possessing better discrimination against off-target sites with non-canonical NAG and NGA PAMs and/or mismatched spacers. We also find that two smaller-size Cas9 orthologues, *Streptococcus thermophilus* Cas9 (St1Cas9) and *Staphylococcus aureus* Cas9 (SaCas9), function efficiently in the bacterial selection systems and in human cells, suggesting that our engineering strategies could be extended to Cas9s from other species. Our findings provide broadly useful SpCas9 variants and, more importantly, establish the feasibility of engineering a wide range of Cas9s with altered and improved PAM specificities.

CRISPR–Cas9 nucleases enable efficient genome editing in a wide variety of organisms and cell types^{1,2}. Target site recognition by Cas9 is programmed by a chimaeric single guide RNA (sgRNA) that encodes a sequence complementary to a target protospacer⁵, but also requires recognition of a short neighbouring PAM^{3–6}. SpCas9, the most robust and widely used Cas9 to date, primarily recognizes NGG PAMs and is consequently restricted to sites that contain this motif^{5,8}. It can therefore be challenging to implement genome editing applications that require precision, such as homology-directed repair, which is most efficient when DSBs are placed within 10–20 base pairs of a desired alteration^{9–11}; the introduction of variable-length insertion or deletion (indel) mutations into small size genetic elements such as microRNAs, splice sites, short open reading frames, or transcription factor binding sites by non-homologous end-joining; and allele-specific editing, where PAM recognition might be exploited to differentiate alleles.

One potential solution to address targeting range limitations would be to engineer Cas9 variants with novel PAM specificities. A previous attempt to alter SpCas9 PAM specificity mutated R1333 and R1335 residues that contact the guanine nucleotides at the second and third PAM positions; however, the R1333Q/R1335Q variant failed to cleave a site harbouring the expected NAA PAM *in vitro*¹². Using a human

U2OS-cell-based enhanced green fluorescent protein (EGFP) reporter gene disruption assay in which nuclease-induced indels lead to loss of fluorescence^{13,14}, we confirmed that an R1333Q/R1335Q SpCas9 variant failed to efficiently cleave target sites with NAA PAMs (Fig. 1a). Additionally, we found that single R1333Q and R1335Q variants each failed to efficiently cleave target sites containing the expected NAG and NGA PAMs, respectively (Fig. 1a), suggesting that re-engineering PAM specificity might require additional mutations.

To identify such mutations, we adapted a bacterial selection system (hereafter referred to as the positive selection) previously used to study properties of homing endonucleases^{15,16}. In our adaptation of this system, survival is enabled by Cas9-mediated cleavage of a selection plasmid encoding an inducible toxic gene (Fig. 1b, Extended Data Fig. 1a). We mutagenized the PAM-interacting domains of wild-type and R1335Q SpCas9 and performed selections against an NGA PAM target site (Extended Data Fig. 1b, Methods). Sequences of surviving clones from both libraries revealed the most frequent substitutions were D1135V/Y/N/E, R1335Q, and T1337R (Extended Data Fig. 2a). After testing all combinations of these mutations using the human cell-based EGFP disruption assay, two variants were chosen for further characterization because they possessed the greatest discrimination between NGA and NGG PAMs: D1135V/R1335Q/T1337R and D1135E/R1335Q/T1337R (hereafter referred to as the VQR and EQR variants, respectively) (Fig. 1c).

To define the global PAM specificity profiles of these SpCas9 variants, we used a bacterial-based negative selection system (Fig. 1d, Extended Data Fig. 3a) similar to other methods previously used to identify PAM preferences of Cas9 (refs 8, 17). In this site-depletion assay, a library of plasmids bearing 6 randomized base pairs adjacent to a protospacer is tested for cleavage by Cas9 in *Escherichia coli* (Extended Data Fig. 3b). Plasmids with PAM sequences refractory to Cas9 enable cell survival due to the presence of an antibiotic resistance gene, whereas plasmids bearing targetable PAMs are depleted from the library (Fig. 1d, Extended Data Fig. 3b). Sequencing the uncleaved population of plasmids enables the calculation of a post-selection PAM depletion value (PPDV), an estimate of Cas9 activity against those PAMs (post-selection frequency relative to the pre-selection frequency). Site-depletion data obtained with catalytically inactive Cas9 (dCas9) on two randomized PAM libraries (each with a different protospacer) enabled us to define what represents a statistically significant change in PPDV for any given PAM or group of PAMs (Extended Data Fig. 3c, d), and PPDVs observed for wild-type SpCas9 recapitulated its previously described profile of targetable PAMs⁸ (Fig. 1e).

Using the site-depletion assay, we obtained PAM specificity profiles for the VQR and EQR variants. The VQR variant strongly depleted sites bearing NGAN and NGCG PAMs, while the EQR variant seemed more specific for an NGAG PAM (Fig. 1f). Human cell

¹Molecular Pathology Unit & Center for Cancer Research, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA. ²Center for Computational and Integrative Biology, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA. ³Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm SE-171 77, Sweden. ⁵Cardiovascular Research Center, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA. ⁶Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁷Broad Institute, Cambridge, Massachusetts 02142, USA. ⁸Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁹Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115, USA.

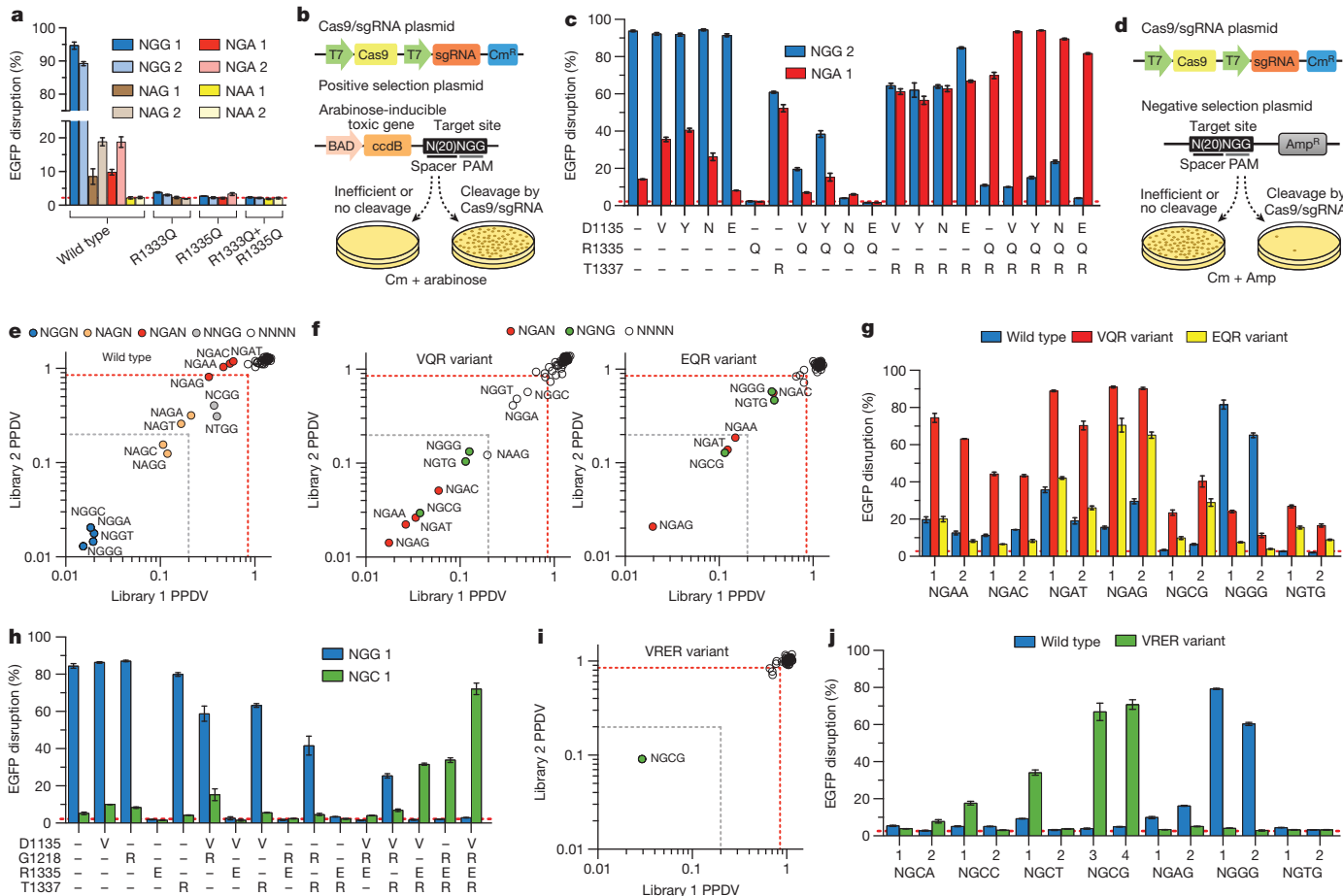


Figure 1 | Evolution and characterization of SpCas9 variants with altered PAM specificities. **a**, Activity of wild-type and mutant SpCas9s assessed via 2OS human cell-based EGFP disruption. Frequencies were quantified by flow cytometry; error bars represent s.e.m., $n = 3$; mean level of background EGFP loss represented by the dashed red line for this and subsequent panels (**c**, **g**, **h** and **j**). **b**, Schematic of the positive selection assay (see also Extended Data Fig. 1). **c**, Combinatorial assembly and testing of mutations obtained from the positive selection for SpCas9 variants that can cleave a target site containing an NGA PAM, using the human cell EGFP disruption assay. **d**, Schematic of the negative selection assay, adapted to profile Cas9 PAM specificity by generating a library of plasmids that contain a randomized sequence adjacent to the 3' end of the protospacer (see also Extended Data Fig. 3a, b). **e**, Scatterplot of the post-selection PAM depletion values (PPDVs) of

EGFP disruption experiments paralleled these results, with the VQR variant robustly cleaving sites bearing NGAN PAMs (with relative efficiencies NGAG > NGAT = NGAA > NGAC), and also sites bearing NGNG PAMs with generally lower efficiencies (Fig. 1g). Similarly, the EQR variant preferred NGAG to the other NGAN and NGNG PAMs in human cells, again at lower activities than with the VQR variant (Fig. 1g). The activities of the VQR and EQR variants in human cells therefore recapitulated what was observed with the bacterial site-depletion assay and suggested that PPDVs of 0.2 (fivefold depletion) provide a reasonable predictive threshold for activity in human cells (Extended Data Fig. 4).

We next sought to extend the generalizability of our engineering strategy by identifying SpCas9 variants capable of recognizing an NGC PAM. Selections using libraries bearing pre-existing R1335E/T1337R and R1335T/T1337R substitutions (Methods) yielded surviving colonies harbouring a variety of additional mutations (Extended Data Fig. 2b). Testing all possible combinations of the most common mutations using the EGFp disruption assay established that the quadruple mutant VRER variant (D1135V/G1218R/R1335E/T1337R) displayed

wild-type SpCas9 with two randomized PAM libraries (each with a different protospacer). PAMs are plotted by their second/third/fourth positions. The red dashed line indicates statistically significant depletion (obtained from a dCas9 control experiment, see Extended Data Fig. 3c), and the grey dashed line represents fivefold depletion (PPDV of 0.2). **f**, PPDV scatterplots for the VQR and EQR variants. **g**, EGFP disruption frequencies for wild-type, VQR, and EQR SpCas9 on sites with NGAN and NGNG PAMs. **h**, Combinatorial assembly and testing of mutations obtained from the positive selection for SpCas9 variants that can cleave a target site containing an NGC PAM, using the human cell EGFP disruption assay. **i**, PPDV scatterplot for the VRER variant. **j**, EGFP disruption frequencies for wild-type and VRER SpCas9 on sites with NGCN and NGNG PAMs.

the highest activity on an NGCG PAM and minimal activity on an NGG PAM (Fig. 1h). Analysis of the VRER variant using the site-depletion assay revealed it to be highly specific for NGCG PAMs (Fig. 1i). Consistent with this result, EGFP disruption assays revealed efficient cleavage of sites with NGCG PAMs, and inconsistent or low activity against NGCH and NGDG PAMs (Fig. 1j). Notably, the mutations critical for altering the specificity of SpCas9 are spatially oriented near the PAM (Extended Data Fig. 5a), and the nature and effect of the mutations imply that they are most likely gain of function (Extended Data Fig. 5b). For example, the T1337R mutation seems to confer a preference for a fourth PAM base, especially in the case of the VRER variant.

To demonstrate directly that the SpCas9 variants broaden the targeting range of SpCas9, we tested their activities against endogenous genes in zebrafish embryos and human cells. In zebrafish embryos, the VQR variant efficiently modified sites bearing NGAG PAMs (range of 20 to 43%, Fig. 2a) with the indels originating at the predicted cleavage sites (Extended Data Fig. 6). In human cells, the VQR variant robustly modified endogenous sites that harboured NGA PAMs (again, with a

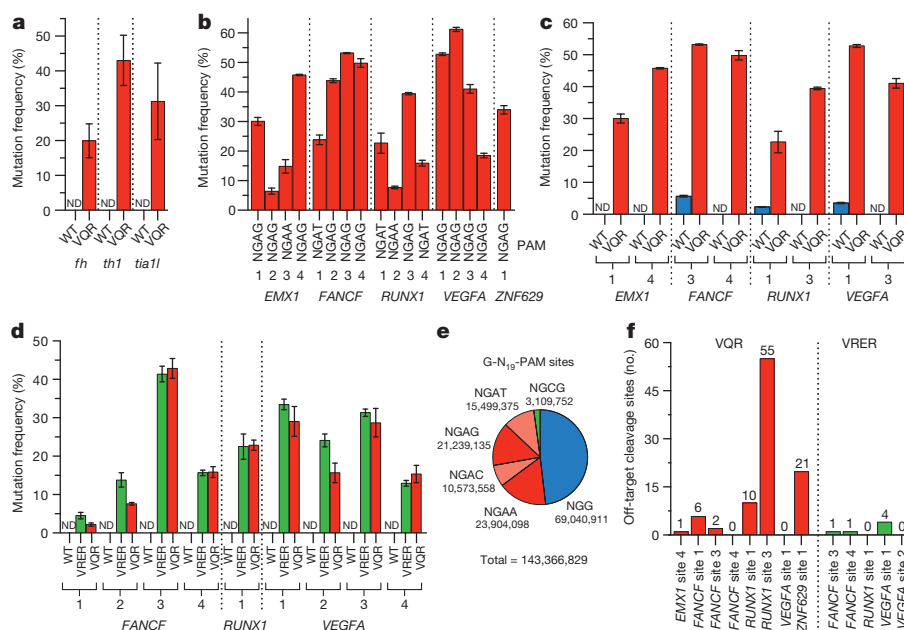


Figure 2 | SpCas9 PAM variants robustly modify endogenous sites in zebrafish embryos and human cells. **a**, Mutagenesis frequencies in zebrafish embryos induced by wild-type or VQR SpCas9 at endogenous gene sites bearing NGAG PAMs. Mutation frequencies were determined using the T7E1 assay; ND, not detectable by T7E1; error bars represent s.e.m., $n = 5$ to 9 embryos. **b**, Endogenous human gene disruption activity of the VQR variant quantified by T7E1 assay. Error bars represent s.e.m., $n = 3$. **c**, Endogenous human gene disruption activity of wild-type SpCas9 against NGA PAM sites quantified by T7E1 assay, where VQR data are re-presented from panel **b** for

ease of comparison. Error bars represent s.e.m., $n = 3$. **d**, Mutation frequencies of wild-type, VRER, and VQR SpCas9 at endogenous human gene sites containing NGCG PAMs quantified by T7E1 assay; error bars represent s.e.m., $n = 3$. **e**, Representation of the number of sites in the human genome with 20-nucleotide spacers potentially targetable by wild-type, VQR, and VRER SpCas9. The 5'-G is included for expression from a U6 promoter. **f**, Number of off-target cleavage sites identified by GUIDE-seq for the VQR and VRER variants using sgRNAs from panels **b** and **d**.

preference for NGAG > NGAT = NGAA, range of 6 to 53%) (Fig. 2b, Extended Data Fig. 7a). Importantly, wild-type SpCas9 was unable to robustly alter NGA PAM sites in zebrafish and human cells (Fig. 2a, c), yet was able to efficiently modify neighbouring sites bearing NGG PAMs in human cells (Extended Data Fig. 7b). When examining VRER variant activity at endogenous human sites with NGCG PAMs, we also observed robust disruption frequencies (range of 5 to 36%) (Fig. 2d). Consistent with the site-depletion data (Fig. 1e, f), the VQR variant also altered NGCG PAM sites while wild-type SpCas9 was unable to do so (Fig. 2d). Taken together, these results demonstrate that the VQR and VRER variants enable modification of previously inaccessible sites in zebrafish embryos and human cells, and computational analysis of the reference human genome reveals that they double the targeting potential of SpCas9 (Fig. 2e). To identify target sites for the engineered variants, we have developed a web-based tool called CasBLASTR (<http://www.CasBLASTR.org>).

To determine the genome-wide specificity of the VQR and VRER SpCas9 nucleases, we used the recently described GUIDE-seq method⁷ to profile off-target cleavage events in human cells. The total number of detectable off-target DSBs induced by the SpCas9 variants in human cells (Fig. 2f) are comparable to (or, in the case of the VRER variant, perhaps less than) what has been previously observed with wild-type SpCas9 (ref. 7). The off-target sites observed generally possess the expected PAM sequences predicted by our site-depletion experiments (compare Figs 1f, i to Extended Data Fig. 8), and the mismatches observed in the off-target sites of the variants are similar to the profiles previously observed with wild-type SpCas9 for sgRNAs targeted to non-repetitive sequences⁷. The stringent genome-wide specificity observed with the VRER variant might result from its extension of the PAM by 1 base pair, and perhaps from the relative depletion of NGCG PAMs in the human genome (Fig. 2e)¹⁸.

Previous studies have shown that imperfect PAM recognition by SpCas9 can lead to recognition of non-canonical PAMs^{7,8,19–21}. While

engineering the VQR variant, we noticed that a D1135E mutant seemed to discriminate between NGG and NGA PAMs better than wild-type SpCas9 (Fig. 1c). Using the site-depletion assay to assess the D1135E variant, we observed a decrease in activity against non-canonical NAG, NGA, and NNGG PAMs relative to wild-type SpCas9, with this effect being more prominent for one protospacer (Fig. 3a). Improved PAM specificity was also observed in human cell EGFP disruption assays, where NAG and NGA PAM sites were less efficiently cleaved by D1135E compared to wild-type SpCas9 (Fig. 3b, mean fold decrease in activity of 1.94). Importantly, wild-type and D1135E SpCas9 had comparable activities against canonical NGG PAM sites when targeted to the EGFP reporter or endogenous human gene sites (mean fold decrease in activity of 1.04) (Fig. 3b and Extended Data Fig. 9a, respectively). It is unlikely that the enhanced specificity of the D1135E variant is the result of protein destabilization, because titration experiments revealed no substantial differences in activity compared with wild-type SpCas9 (Extended Data Fig. 9b).

To more directly assess the effect of D1135E on off-target effects, we examined the mutation rates induced by wild-type and D1135E SpCas9 on 25 previously known off-target sites of three sgRNAs^{7,14,19}. Deep-sequencing revealed that D1135E improved specificity for 19 of the 22 off-target sites with mutation frequencies above background indel rates, when compared to the relative mutation frequencies observed at the on-target sites (Fig. 3c, Extended Data Fig. 9c). Interestingly, the gains in specificity with D1135E are not restricted to sites with non-canonical PAMs. To more thoroughly assess the improvements in specificity associated with the D1135E variant, we performed GUIDE-seq using three different sgRNAs and observed a generalized improvement in genome-wide specificity relative to wild-type SpCas9 (Fig. 3d, Extended Data Fig. 9d–f). Collectively, these results show that the D1135E substitution increases the specificity of SpCas9.

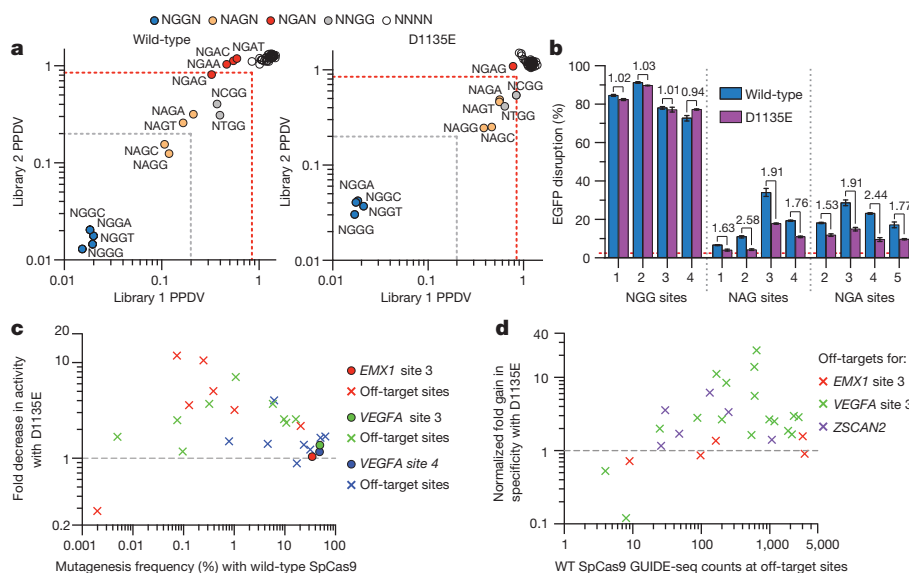


Figure 3 | A D1135E mutation improves the PAM recognition and spacer specificity of SpCas9. **a**, PPDV scatterplots for wild-type and D1135E SpCas9 for the two randomized PAM libraries. PAMs are plotted by their second/third/fourth positions, and wild-type data are the same as shown in Fig. 1e for ease of comparison. The red dashed line indicates PAMs that are statistically significantly depleted (see Extended Data Fig. 3c), and the grey dashed line indicates fivefold depletion (PPDV of 0.2). **b**, EGFP disruption activities of wild-type and D1135E SpCas9 on sites that contain canonical and non-canonical PAMs in human cells. Disruption frequencies were quantified by

flow cytometry; mean background level of EGFP loss represented by the dashed red line; error bars represent s.e.m., $n = 3$; fold change in activity is shown. **c**, Summary of targeted deep-sequencing data demonstrating specificity gains at off-target sites when using D1135E (see also Extended Data Fig. 9c). **d**, Summary of GUIDE-seq detected changes in specificity between wild-type and D1135E at off-target sites (see also Extended Data Fig. 9f). Estimated fold gain in specificity at sites without read counts for D1135E are not plotted (see Extended Data Fig. 9f).

The many Cas9 orthologues from other bacteria make attractive candidates for characterizing and engineering Cas9s with novel PAM specificities^{22,23}. To explore this, we determined whether two smaller-size orthologues, *Streptococcus thermophilus* Cas9 from the CRISPR1 locus (St1Cas9)^{24,25} and *Staphylococcus aureus* (SaCas9)²³ could function in the bacterial selection assays. Although the PAM

of St1Cas9 has previously been characterized as NNAGAA^{17,22,24,25}, our attempts to bioinformatically derive the SaCas9 PAM using a previously described approach²² failed to yield a consensus sequence. Therefore, we used the site-depletion assay to determine the PAM for SaCas9 and, as a positive control, St1Cas9. For St1Cas9, we identified two novel PAMs in addition to six PAMs that had been

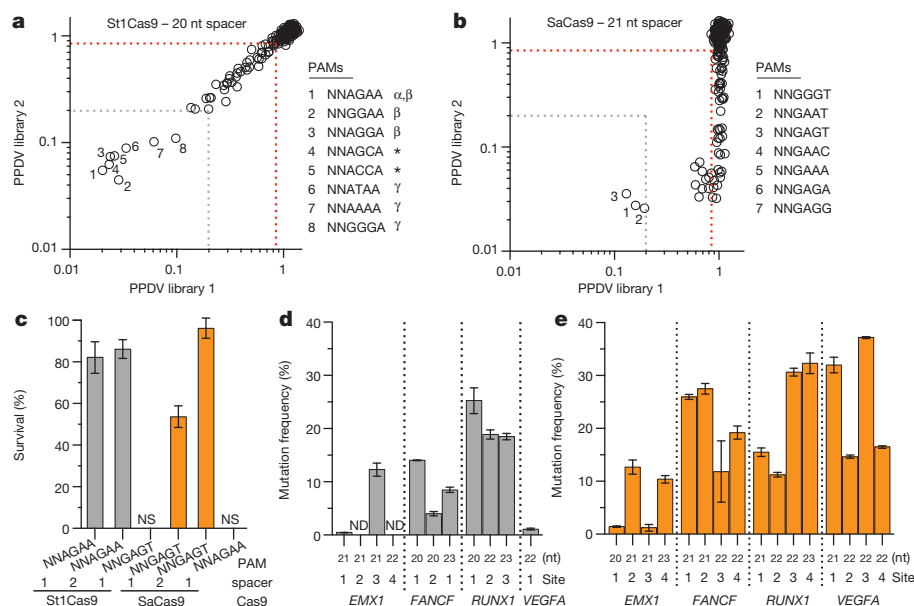


Figure 4 | Characterization of St1Cas9 and SaCas9 in bacteria and human cells. **a**, **b**, PPDV scatterplots for St1Cas9 (**a**) and SaCas9 (**b**), with PAMs plotted by their third/fourth/fifth/sixth positions. The red dashed line indicates PAMs that are significantly depleted (Extended Data Fig. 3c), and the grey dashed line represents fivefold depletion (PPDV of 0.2); α , PAM previously predicted by a bioinformatic approach²⁵; β , PAMs previously identified under stringent experimental conditions¹⁷; γ , novel PAMs discovered in this study;

γ , PAMs previously identified under moderate experimental conditions¹⁷. **c**, Survival percentages of St1Cas9 and SaCas9 in the bacterial positive selection when challenged with selection plasmids that harbour different spacer sequences and PAMs. NS, no survival. **d**, **e**, Mutation frequencies of St1Cas9 (**d**) and SaCas9 (**e**) quantified by T7E1 assay at sites in four endogenous human genes. Error bars represent s.e.m., $n = 3$; ND, not detectable by T7E1; nt, nucleotide.

previously described^{17,22,25} (Fig. 4a, Extended Data Fig. 10a, b). For SaCas9, only three PAMs were depleted more than fivefold in all experiments (NNGGGT, NNGAAT, NNGAGT, Fig. 4b), although additional PAMs were targetable when using the second protospacer library (Extended Data Fig. 10c, d). These results are consistent with a recent definition of SaCas9 PAM specificity²³. We also found that St1Cas9 and SaCas9 can function efficiently in the bacterial positive selection system (Fig. 4c), suggesting that their PAM specificities could potentially be modified by mutagenesis and selection.

Because not all Cas9 orthologues function efficiently outside of their native context^{17,23}, we tested whether St1Cas9 and SaCas9 can modify sites in human cells. St1Cas9 has been previously shown to function as a nuclease in human cells but only on four sites^{17,23,26}, and a recently published manuscript assessed SaCas9 activity²³. In EGFP disruption experiments, St1Cas9 displayed high activity at three of five target sites and SaCas9 efficiently targeted eight sites (Extended Data Fig. 10e). No obvious correlation between activity and length of spacer was observed (Extended Data Fig. 10e, f). When examining activity on endogenous loci, St1Cas9 efficiently targeted 7 out of 11 sites (1 to 25% disruption; Fig. 4d), SaCas9 displayed more robust activity at 16 sites (1% to 37%; Fig. 4e), and again no distinct spacer length requirement was observed (Extended Data Fig. 10g). Collectively, these results demonstrate that St1Cas9 and SaCas9 function in human cells, making them attractive candidates for engineering additional variants with novel PAM specificities.

The VQR and VRER variants engineered in this study enhance the opportunities to utilize the CRISPR-Cas9 platform to practice efficient homology-directed repair, to generate non-homologous end-joining-mediated indels in small genetic elements, and to exploit the requirement for a PAM to distinguish between different alleles in the same cell. Importantly, the VQR, VRER, and D1135E variants all have similar (or better) genome-wide specificities compared to wild-type SpCas9. These variants can be rapidly incorporated into existing and widely used SpCas9 vectors by simple site-directed mutagenesis, and we expect that the variants should also work with other previously described improvements to the SpCas9 platform (for example, truncated sgRNAs^{7,27}, SpCas9 nickases^{20,28}, or dimeric FokI-dCas9 fusions^{29,30}). Collectively, our results establish engineering PAM recognition and characterization of additional Cas9 orthologues (as previously described)^{17,22,23} as complementary approaches to provide researchers with an expanded repertoire of genome-editing reagents, while also demonstrating the feasibility of engineering Cas9 nucleases with useful new properties.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 March 2015; accepted 1 June 2015.

Published online 22 June 2015.

- Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnol.* **32**, 347–355 (2014).
- Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
- Mojica, F. J., Díez-Villasenor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
- Shah, S. A., Erdmann, S., Mojica, F. J. & Garrett, R. A. Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biol.* **10**, 891–899 (2013).
- Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
- Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
- Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnol.* **33**, 187–197 (2015).

- Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnol.* **31**, 233–239 (2013).
- Yang, L. et al. Optimization of scarless human stem cell genome editing. *Nucleic Acids Res.* **41**, 9049–9061 (2013).
- Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J. A. & Jasin, M. Gene conversion tracts from double-strand break repair in mammalian cells. *Mol. Cell Biol.* **18**, 93–101 (1998).
- Findlay, G. M., Boyle, E. A., Hause, R. J., Klein, J. C. & Shendure, J. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* **513**, 120–123 (2014).
- Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573 (2014).
- Reyon, D. et al. FLASH assembly of TALENs for high-throughput genome editing. *Nature Biotechnol.* **30**, 460–465 (2012).
- Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnol.* **31**, 822–826 (2013).
- Chen, Z. & Zhao, H. A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Res.* **33**, e154 (2005).
- Doyon, J. B., Pattanayak, V., Meyer, C. B. & Liu, D. R. Directed evolution and substrate specificity profile of homing endonuclease I-SceI. *J. Am. Chem. Soc.* **128**, 2477–2484 (2006).
- Esvelt, K. M. et al. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature Methods* **10**, 1116–1121 (2013).
- Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- Hsu, P. D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnol.* **31**, 827–832 (2013).
- Mali, P. et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnol.* **31**, 833–838 (2013).
- Zhang, Y. et al. Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. *Sci. Rep.* **4**, 5405 (2014).
- Fonfara, I. et al. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res.* **42**, 2577–2590 (2014).
- Ran, F. A. et al. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature*, (2015).
- Deveau, H. et al. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400 (2008).
- Horvath, P. et al. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1401–1412 (2008).
- Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
- Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnol.* **32**, 279–284 (2014).
- Ran, F. A. et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).
- Guilinger, J. P., Thompson, D. B. & Liu, D. R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nature Biotechnol.* **32**, 577–582 (2014).
- Tsai, S. Q. et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature Biotechnol.* **32**, 569–576 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Edgell for providing the bacterial strain and plasmids related to the bacterial selection; J. Angstman and V. Pattanayak for discussion and comments on the manuscript. This work was supported by a National Institutes of Health (NIH) Director's Pioneer Award (DP1 GM105378) and NIH R01 GM107427 to J.K.J., NIH R01 GM088040 to J.K.J. and R.T.P., The Jim and Ann Orr Research Scholar Award (to J.K.J.), and a National Sciences and Engineering Research Council of Canada Postdoctoral Fellowship (to B.P.K.).

Author Contributions B.P.K., M.S.P., S.Q.T. and N.T.N. performed all bacterial and human cell-based experiments. A.P.W.G. and Z.L. performed all zebrafish experiments. S.Q.T., V.T., Z.Z. and M.J.A. analysed the site-depletion, targeted deep-sequencing, and GUIDE-seq data. B.P.K., R.T.P., J.-R.J.Y. and J.K.J. directed the research and interpreted experiments. B.P.K. and J.K.J. wrote the manuscript with input from all the authors.

Author Information All new reagents described in this work have been deposited with the non-profit plasmid distribution service Addgene (<http://www.addgene.org/crispr-cas>). A web-tool to design sgRNA sites for the engineered variants and orthogonal Cas9 nucleases described in this study can be found at <http://www.CasBLAST.org>. The sequences generated in this study have been deposited in the Sequences Read Archive under accession number SRP058629. Reprints and permissions information is available at www.nature.com/reprints. Readers are welcome to comment on the online version of the paper. The authors declare competing financial interests: details are available in the online version of the paper. Correspondence and requests for materials should be addressed to J.K.J. (jjoung@mgh.harvard.edu).

METHODS

No statistical methods were used to predetermine sample size, and the investigators were not blinded to allocation during experiments and outcome assessment.

Plasmids and oligonucleotides. DNA sequences for parent constructs used in this study can be found in Supplementary Information. Sequences of oligonucleotides used to generate the positive selection plasmids, negative selection plasmids, and site-depletion libraries are available in Supplementary Table 1. Sequences of all sgRNA targets in this study are available in Supplementary Table 2. Point mutations in Cas9 were generated by PCR. For cloning purposes, please note the low copy number origins of these plasmids. All new plasmids described in this study will be deposited with the non-profit plasmid distribution service Addgene: <http://www.addgene.org/crispr-cas>.

Bacterial Cas9/sgRNA expression plasmids were constructed with two T7 promoters to separately express Cas9 and the sgRNA. These plasmids encode human codon optimized versions of Cas9 for *S. pyogenes* (BPK764, SpCas9 sequence subcloned from JDS246; ref. 14), *S. thermophilus* Cas9 from CRISPR locus 1 (MSP1673, St1Cas9 sequence modified from previous published description¹⁷), and *S. aureus* (BPK2101, SaCas9 sequence codon optimized from Uniprot J7RUA5). Previously described sgRNA sequences were used for SpCas9 (refs 31, 32) and St1Cas9 (ref. 17), while the SaCas9 sgRNA sequence was determined by searching the European Nucleotide Archive sequence HE980450 for crRNA repeats using CRISPRfinder (<http://crispr.u-psud.fr/Server/>) and identifying the tracrRNA using a bioinformatic approach similar to one previously described³³. Annealed oligonucleotides to complete the spacer complementarity region of the sgRNA were ligated into BsaI-cut BPK764 and BPK2101, or BspMI-cut MSP1673 (append 5'-ATAG to the spacer to generate the top oligo and append 5'-AAAC to the reverse complement of the spacer sequence to generate the bottom oligo). A 5'-GG dinucleotide was included on all bacterial plasmid sgRNAs for proper expression from the T7 promoter.

Residues 1097–1368 of SpCas9 were randomly mutagenized using Mutazyme II (Agilent Technologies) at a rate of ~5.2 substitutions/kilobase to generate mutagenized PAM-interacting domain libraries. For NGA PAM selections, wild-type SpCas9 and R1335Q were used as templates for mutagenesis. For NGC PAM selections, we first designed Cas9 mutants bearing amino acid substitutions of R1335 that might be expected to interact with a cytosine (D, E, S, or T) and found no activity on an NGC PAM site using the positive selection system (data not shown). We then randomly mutagenized the PAM-interacting domain of each of these singly substituted variants but still failed to obtain surviving colonies in positive selections (data not shown). Because the T1337R mutation had increased the activities of our VQR and EQR variants, we combined this mutation with R1335 substitutions of A, D, E, S, T, or V, and again randomly mutagenized their PAM-interacting domains. Selections using two of these six mutagenized libraries (bearing pre-existing R1335E/T1337R and R1335T/T1337R substitutions) yielded surviving colonies harbouring a variety of additional mutations (Extended Data Fig. 2b). The theoretical complexity of each PAM-interacting domain library was estimated to be greater than 10^7 clones based on the number of transformants obtained. Positive and negative selection plasmids were generated by ligating annealed target site oligonucleotides into XbaI/SphI or EcoRI/SphI cut p11-lacY-wtx1¹⁵, respectively.

Two randomized PAM libraries (each with a different protospacer sequence) were constructed using Klenow(-exo) to fill-in the bottom strand of oligonucleotides that contained six randomized nucleotides directly adjacent to the 3' end of the protospacer (see Supplementary Table 1). The double-stranded product was cut with EcoRI to leave EcoRI/SphI ends for ligation into cut p11-lacY-wtx1. The theoretical complexity of each randomized PAM library was estimated to be greater than 10^6 based on the number of transformants obtained.

SpCas9 and variants were expressed in human cells from vectors derived from JDS246 (ref. 14). For St1Cas9 and SaCas9, the Cas9 ORFs from MSP1673 and BPK2101 were subcloned into a CAG promoter vector to generate MSP1594 and BPK2139, respectively. Plasmids for U6 expression of sgRNAs (into which desired spacer oligonucleotides can be cloned) were generated using the sgRNA sequences described above for the SpCas9 sgRNA (BPK1520), the St1Cas9 sgRNA (BPK2301), and the SaCas9 sgRNA (VVT1). Annealed oligonucleotides to complete the spacer complementarity region of the sgRNA were ligated into the BsmBI overhangs of these vectors (append 5'-CACC to the spacer to generate the top oligo and append 5'-AAAC to the reverse complement of the spacer sequence to generate the bottom oligo). A 5'-G of target spacer sequences was included when designing human cell sgRNAs, for proper expression from the U6 promoter (and thus included in the calculation in Fig. 2e).

Bacterial-based positive selection assay for evolving SpCas9 variants. Competent *E. coli* BW25141(λ .DE3)³⁴ containing a positive selection plasmid (with embedded target site) were transformed with Cas9/sgRNA-encoding

plasmids. Following a 60 min recovery in SOB media, transformations were plated on LB plates containing either chloramphenicol (non-selective) or chloramphenicol + 10 mM arabinose (selective). Cleavage of the positive selection plasmid was estimated by calculating the survival frequency: colonies on selective plates/colonies on non-selective plates (see also Extended Data Fig. 1).

To select for SpCas9 variants that can target novel PAMs, PAM-interacting-domain mutagenized Cas9/sgRNA plasmid libraries were electroporated into *E. coli* BW25141(λ .DE3) cells containing a positive selection plasmid that encodes a target site and PAM of interest. Generally ~50,000 clones were screened to obtain between 50 and 100 survivors. The PAM-interacting domains of surviving clones were subcloned into fresh backbone plasmid and re-tested in the positive selection. Clones that had greater than 10% survival in this secondary screen for activity were sequenced. Mutations observed in the sequenced clones were chosen for further assessment based on their frequency in surviving clones, type of substitution, proximity to the PAM bases in the SpCas9–sgRNA crystal structure (PDB:4UN3)¹², and (in some cases) activities in a human cell-based EGFP disruption assay.

Bacterial-based site-depletion assay for profiling Cas9 PAM specificities. Competent *E. coli* BW25141(λ .DE3) containing a Cas9/sgRNA expression plasmid were transformed with negative selection plasmids harbouring cleavable or non-cleavable target sites. Following a 60 min recovery in SOB media, transformations were plated on LB plates containing chloramphenicol + carbenicillin. Cleavage of the negative selection plasmid was estimated by calculating the colony forming units per μ g of DNA transformed (see also Extended Data Fig. 3).

The negative selection was adapted to determine PAM specificity profiles of Cas9 nucleases by electroporating each randomized PAM library into *E. coli* BW25141(λ .DE3) cells harbouring an appropriate Cas9/sgRNA plasmid. Between 80,000 and 100,000 colonies were plated at a low density spread on LB + chloramphenicol + carbenicillin plates. Surviving colonies containing negative selection plasmids refractory to cleavage by Cas9 were harvested and plasmid DNA isolated by maxi-prep (Qiagen). The resulting plasmid library was amplified by PCR using Phusion Hot-start Flex DNA Polymerase (New England BioLabs) followed by an Agencourt Ampure XP clean-up step (Beckman Coulter Genomics). Dual-indexed Tru-seq Illumina deep-sequencing libraries were prepared using the KAPA HTP library preparation kit (KAPA BioSystems) from ~500 ng of clean PCR product for each site-depletion experiment. The Dana-Farber Cancer Institute Molecular Biology Core performed 150-bp paired-end sequencing on an Illumina MiSeq Sequencer.

The raw FASTQ files outputted for each MiSeq run were analysed with a Python program to determine relative PAM depletion. The program (see Supplementary Information) operates as follows: first, a file dialogue is presented to the user from which all FASTQ read files for a given experiment can be selected. For these files, each FASTQ entry is scanned for the fixed spacer region on both strands. If the spacer region is found, then the six variable nucleotides flanking the spacer region are captured and added to a counter. From this set of detected variable regions, the count and frequency of each window of length 2–6 nucleotides at each possible position was tabulated (see Supplementary Table 3 for the 6-nucleotide output). The site-depletion data for both randomized PAM libraries was analysed by calculating the post-selection PAM depletion value (PPDV): the post-selection frequency of a PAM in the selected population divided by the pre-selection library frequency of that PAM. PPDV analyses were performed for each experiment across all possible 2–6 length windows in the 6-bp randomized region. The windows we used to visualize PAM preferences were: the 3-nucleotide window representing the second, third and fourth PAM positions for wild-type and variant SpCas9 experiments, and the 4-nucleotide window representing the third, fourth, fifth and sixth PAM positions for St1Cas9 and SaCas9.

Two significance thresholds for PPDVs were determined based on: (1) a statistical significance threshold based on the distribution of dCas9 versus pre-selection library log read count ratios (see Extended Data Fig. 3c, d), and (2) a biological activity threshold based on an empirical correlation between depletion values and activity in human cells. The statistical threshold was set at 3.36 s.d. from the mean PPDV for dCas9 (equivalent to a relative PPDV of 0.85), corresponding to a normal distribution two-sided *P* value of 0.05 after adjusting for multiple comparisons (that is, *P* = 0.05/64). The biological activity threshold was set at fivefold depletion (equivalent to a PPDV of 0.2) because this level of depletion serves as a reasonable predictor of activity in human cells (see also Extended Data Fig. 4). The 95% confidence intervals in Extended Data Fig. 4 were calculated by dividing the standard deviation of the mean by the square root of the sample size multiplied by 1.96.

Human cell culture and transfection. U2OS cells obtained from our collaborator T. Cathomen (Freiburg) and U2OS.EGFP cells harbouring a single integrated copy of a constitutively expressed EGFP–PEST reporter gene¹³ were cultured in Advanced DMEM media (Life Technologies) supplemented with 10% FBS, 2 mM GlutaMAX (Life Technologies), penicillin/streptomycin, at 37 °C with 5%

CO₂. Additionally, U2OS.EGFP cells were cultured in 400 µg ml⁻¹ of G418. The identity of U2OS and U2OS.EGFP cell lines were validated by STR profiling (ATCC) and deep sequencing, and cells were tested bi-weekly for mycoplasma contamination. Cells were co-transfected with 750 ng of Cas9 plasmid and 250 ng of sgRNA plasmid (unless otherwise noted) using the DN-100 program of a Lonza 4D-nucleofactor according to the manufacturer's protocols. Cas9 plasmid transfected together with an empty U6 promoter plasmid was used as a negative control for spontaneous background EGFP loss for all human cell EGFP disruption experiments, and all endogenous gene disruption experiments (none of which showed detectable activity by T7E1). Target sites for endogenous gene experiments were selected within 200 bp of NGG sites cleavable by wild-type SpCas9 (see Extended Data Fig. 7a and Supplementary Table 2).

Zebrafish care and injections. Zebrafish care and use was approved by the Massachusetts General Hospital Subcommittee on Research Animal Care. Cas9 mRNA was transcribed with PmeI-digested JDS246 (wild-type SpCas9) or MSP469 (VQR variant) using the mMESSAGE mMACHINE T7 ULTRA Kit (Life Technologies) as previously described³². All sgRNAs in this study were prepared according to the cloning-independent sgRNA generation method³⁵. sgRNAs were transcribed by the MEGAscript SP6 Transcription Kit (Life Technologies), purified by RNA Clean & Concentrator-5 (Zymo Research), and eluted with RNase-free water.

sgRNA- and Cas9-encoding mRNA were co-injected into one-cell stage zebrafish embryos. Each embryo was injected with ~2–4.5 nl of solution containing 30 ng µl⁻¹ sgRNA and 300 ng µl⁻¹ Cas9 mRNA. The next day, injected embryos were inspected under a stereoscope for normal morphological development, and genomic DNA was extracted from 5 to 9 embryos.

Human cell EGFP disruption assay. EGFP disruption experiments were performed as previously described¹⁴. Transfected cells were analysed for EGFP expression ~52 h post-transfection using a Fortessa flow cytometer (BD Biosciences). Background EGFP loss was gated at approximately 2.5% for all experiments (graphically represented as a dashed red line).

T7E1 assay, targeted deep-sequencing, and GUIDE-seq to quantify nuclease-induced mutations. T7E1 assays were performed as previously described for human cells¹³ and zebrafish³². For human cells, genomic DNA was extracted from transfected cells ~72 h post-transfection using the Agencourt DNAdvance Genomic DNA Isolation Kit (Beckman Coulter Genomics). Target loci from zebrafish or human cell genomic DNA were amplified using the primers listed in Supplementary Table 1. Roughly 200 ng of purified PCR product was denatured, annealed, and digested with T7E1 (New England BioLabs). Mutagenesis frequencies were quantified using a Qiaxcel capillary electrophoresis instrument (Qiagen), as previously described for human cells¹³ and zebrafish³².

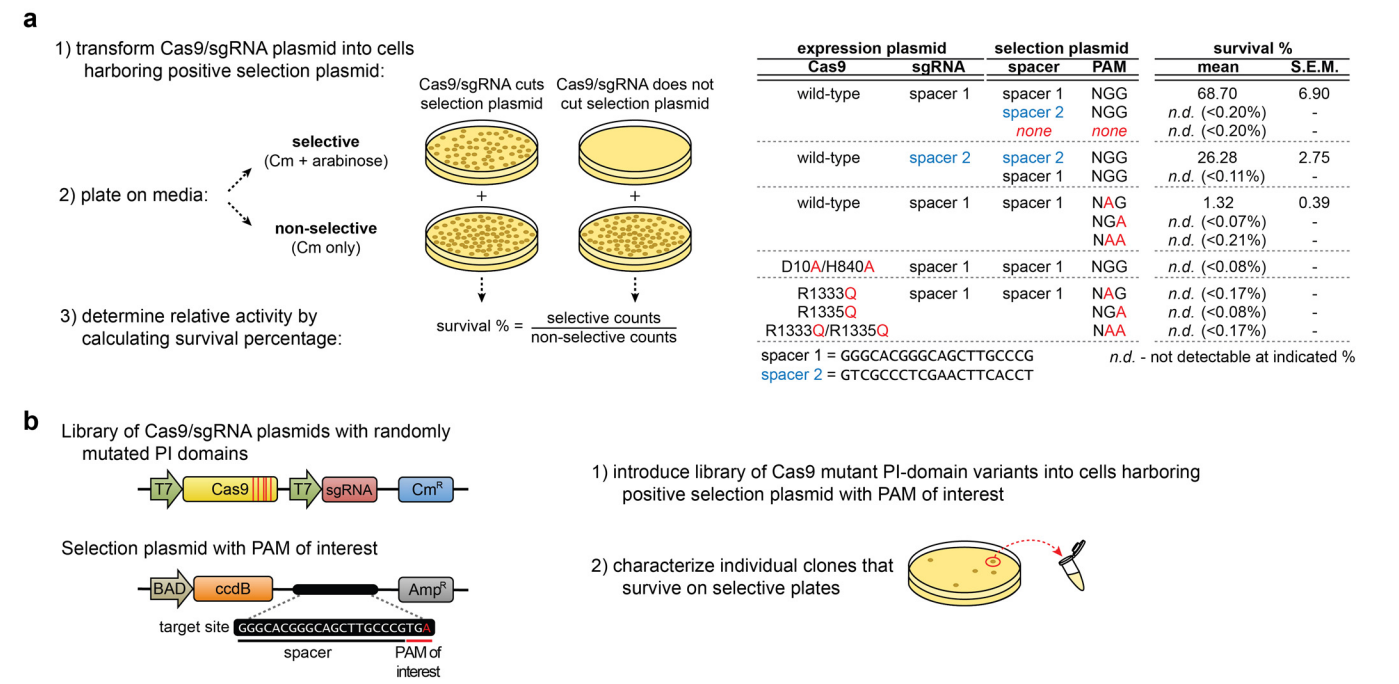
For targeted deep-sequencing, previously characterized on- and off-target sites^{7,14,27} were amplified using Phusion Hot-start Flex with the primers listed in Supplementary Table 1. Genomic loci were amplified for a control condition (empty sgRNA), wild-type, and D1135E SpCas9. An Agencourt Ampure XP clean-up step (Beckman Coulter Genomics) was performed before pooling ~500 ng of DNA from each condition for library preparation. Dual-indexed Tru-seq Illumina deep-sequencing libraries were generated using the KAPA HTP library preparation kit (KAPA BioSystems). The Dana-Farber Cancer

Institute Molecular Biology Core performed 150-bp paired-end sequencing on an Illumina MiSeq Sequencer. Mutation analysis of targeted deep-sequencing data was performed as previously described³⁰. Briefly, Illumina MiSeq paired end read data was mapped to human genome reference GRCh37 using bwa³⁶. High-quality reads (quality score ≥ 30) were assessed for indel mutations that overlapped the target or off-target sites. 1-bp indel mutations were excluded from the analysis unless they occurred within 1-bp of the predicted breakpoint. Changes in activity at on- and off-target sites comparing D1135E versus wild-type SpCas9 were calculated by comparing the indel frequencies from both conditions (for rates above background control amplicon indel levels).

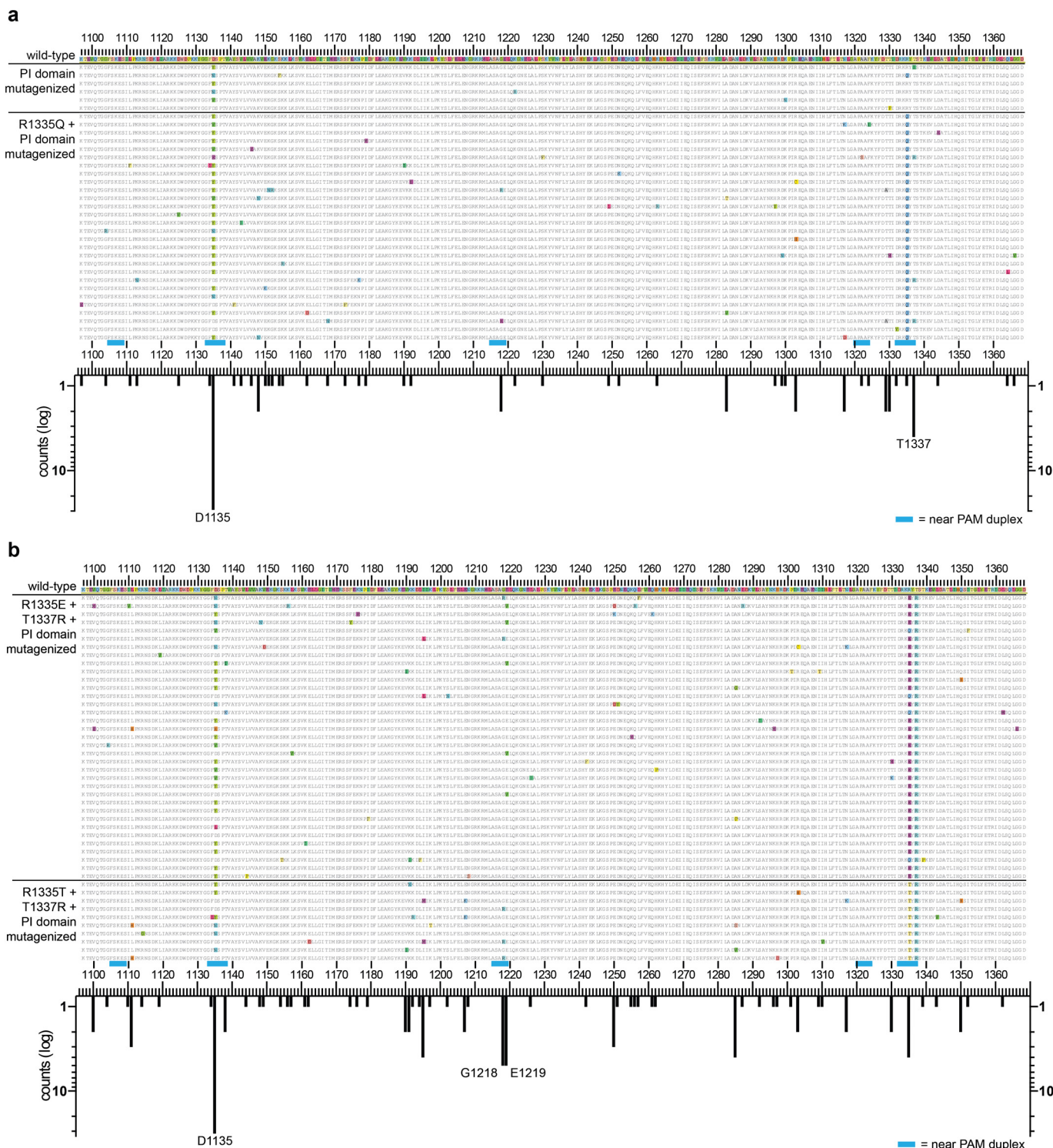
GUIDE-seq experiments were performed as previously described⁷. Briefly, 100 pmol of phosphorylated, phosphorothioate-modified double-stranded oligodeoxynucleotides (dsODNs) were transfected into U2OS cells along with Cas9 and sgRNA expression plasmids, as described above. dsODN-specific amplification, high-throughput sequencing, and mapping were performed to identify genomic intervals containing DSB activity. For wild-type versus D1135E experiments, off-target read counts were normalized to the on-target read counts to correct for sequencing depth differences between samples. The normalized ratios for wild-type and D1135E SpCas9 were then compared to calculate the fold change in activity at off-target sites. To determine whether wild-type and D1135E samples for GUIDE-seq had similar oligo tag integration rates at the intended target site, restriction fragment length polymorphism (RFLP) assays were performed by amplifying the intended target loci with Phusion Hot-Start Flex from 100 ng of genomic DNA (isolated as described above) using primers listed in Supplementary Table 1. Roughly 500 ng of PCR product was digested with 20 U of NdeI (New England BioLabs) for 3 h at 37 °C before clean-up using the Agencourt Ampure XP kit. RFLP results were quantified using a Qiaxcel capillary electrophoresis instrument (Qiagen) to approximate oligo tag integration rates. T7E1 assays were performed for a similar purpose, as described above. For the quantitative comparison of wild-type to D1135E SpCas9, we utilized an alternative sequence consolidation algorithm that is more stringent and less likely to overestimate the number of unique molecularly-indexed GUIDE-seq reads. All sequencing data was corrected for cell-type specific single nucleotide polymorphisms.

Code availability. Custom code written to analyse PAM depletion MiSeq data is shown in the Supplementary Information.

31. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
32. Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnol.* **31**, 227–229 (2013).
33. Chylinski, K., Le Rhun, A. & Charpentier, E. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol.* **10**, 726–737 (2013).
34. Kleinstiver, B. P., Fernandes, A. D., Gloor, G. B. & Edgell, D. R. A unified genetic, computational and experimental framework identifies functionally relevant residues of the homing endonuclease I-Bmol. *Nucleic Acids Res.* **38**, 2411–2427 (2010).
35. Gagnon, J. A. *et al.* Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS ONE* **9**, e98186 (2014).
36. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

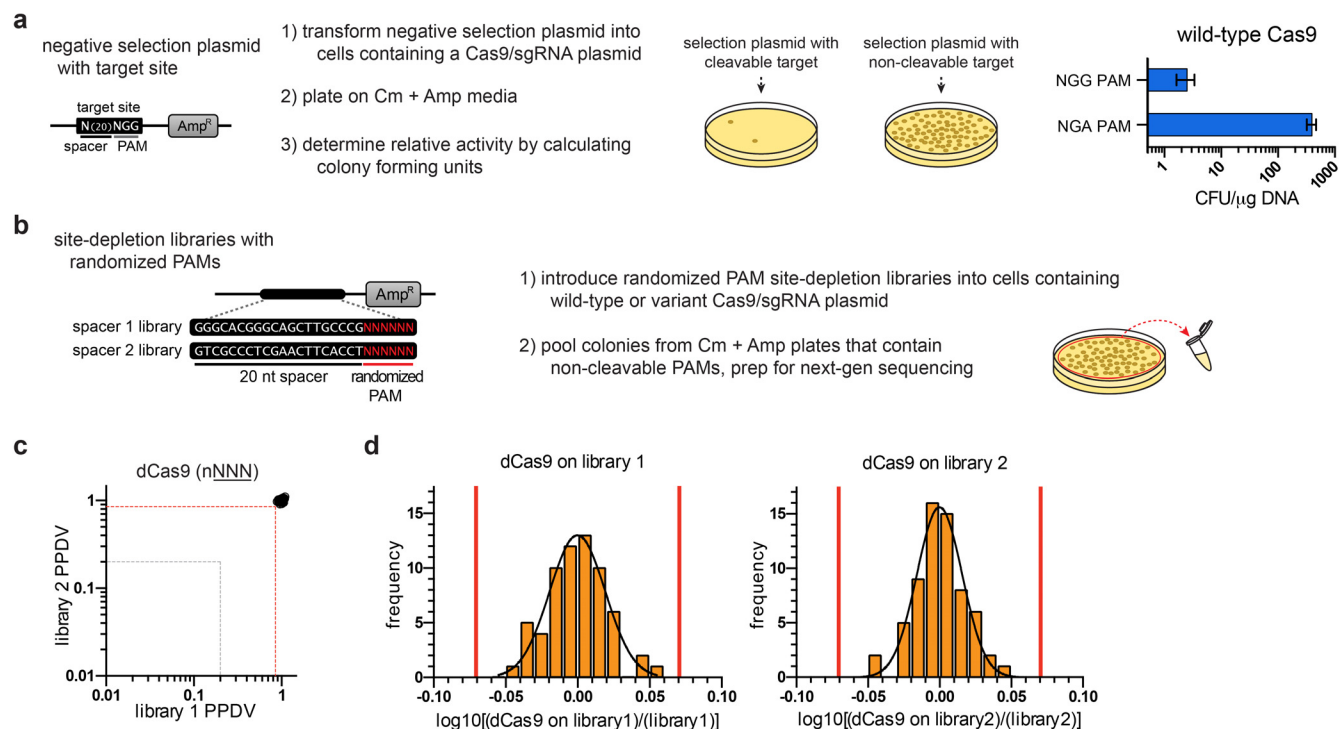


Extended Data Figure 1 | Bacterial-based positive selection used to engineer altered PAM specificity variants of SpCas9. **a**, Expanded schematic of the positive selection from Fig. 1b (left panel), and validation that SpCas9 behaves as expected in the positive selection (right panel). **b**, Schematic of how the positive selection was adapted to select for SpCas9 variants that have altered PAM recognition specificities. A library of SpCas9 clones with randomized PAM-interacting (PI) domains (residues 1097–1368) is challenged by a selection plasmid that harbours an altered PAM. Variants that survive the selection by cleaving the positive selection plasmid are sequenced to determine the mutations that enable altered PAM specificity.



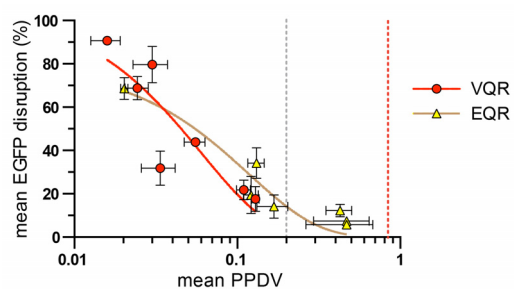
Extended Data Figure 2 | Amino acid sequences of clones that cleave target sites bearing alternate PAMs in the bacterial-based positive selection system. **a**, Sequences of variants that survived >10% when re-tested in the positive selection assay against an NGA PAM site (see Methods). Variants were selected from libraries containing randomly mutagenized PAM-interacting domains (residues 1097–1368) with or without a starting R1335Q mutation. Sequence differences compared with wild-type SpCas9 are highlighted. The histogram below illustrates the number of changes at each position (not counting the

starting R1335Q mutation). **b**, Sequences of variants that survived >10% when re-tested in the positive selection assay against a site containing an NGC PAM. Variants were selected from libraries containing randomly mutagenized PAM-interacting domains (residues 1097–1368) with starter mutation pairs of R1335E/T1337R or R1335T/T1337R. Sequence differences compared with wild-type SpCas9 (shown at the top) are highlighted. The histogram below illustrates the number of changes at each position (not counting starter mutations at R1335 or T1337).

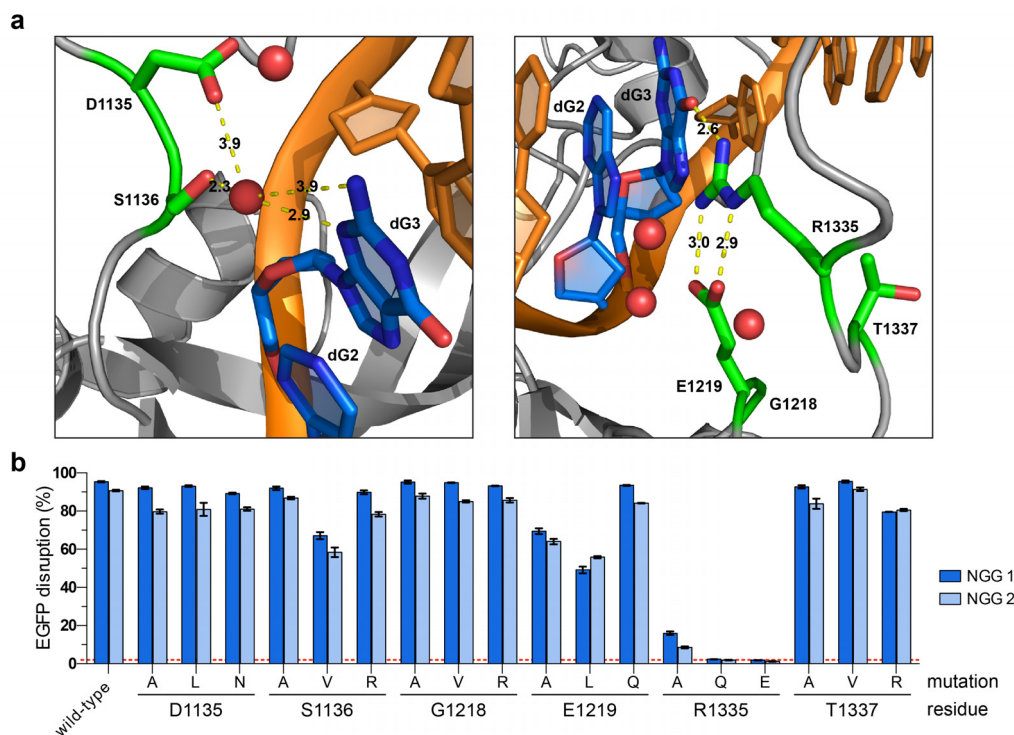


Extended Data Figure 3 | Bacterial cell-based site-depletion assay for profiling the global PAM specificities of Cas9 nucleases. **a**, Expanded schematic illustrating the negative selection from Fig. 1d (left panel), and validation that wild-type SpCas9 behaves as expected in a screen of sites with functional (NGG) and non-functional (NGA) PAMs (right panel). **b**, Schematic of how the negative selection was used as a site-depletion assay to screen for functional PAMs by constructing negative selection plasmid libraries containing 6 randomized base pairs in place of the PAM. Selection plasmids that contain PAMs cleaved by a Cas9/sgRNA of interest are depleted while PAMs that are not cleaved (or poorly cleaved) are retained. The frequencies of

the PAMs following selection are compared to their pre-selection frequencies in the starting libraries to calculate the post-selection PAM depletion value (PPDV). **c**, **d**, A cutoff for statistically significant PPDVs was established by plotting the PPDV of PAMs for catalytically inactive SpCas9 (dCas9) (grouped and plotted by their second/third/fourth positions) for the two randomized PAM libraries (**c**). A threshold of 3.36 standard deviations from the mean PPDV for the two libraries was calculated (red lines in (**d**)), establishing that any PPDV deviation below 0.85 is statistically significant compared to dCas9 treatment (red dashed line in (**c**)). The grey dashed line in (**c**) indicates a fivefold depletion in the assay (PPDV of 0.2).



Extended Data Figure 4 | Concordance between the site-depletion assay and EGFP disruption activity. Data points represent the average EGFP disruption of the two NGAN and NGNG PAM sites for the VQR and EQR variants (Fig. 1g) plotted against the mean PPDV observed for library 1 and 2 (Fig. 1f) for the corresponding PAM. The red dashed line indicates PAMs that are statistically significantly depleted (PPDV of 0.85, see Extended Data Fig. 3c), and the grey dashed line represents fivefold depletion (PPDV of 0.2). Mean values are plotted with the 95% confidence interval.



Extended Data Figure 5 | Structural and functional roles of D1135, G1218, and T1337 in PAM recognition by SpCas9. **a**, Structural representations of the six residues implicated in PAM recognition. The left panel illustrates the proximity of D1135 to S1136, a residue that makes a water-mediated, minor groove contact to the third base position of the PAM¹². The right panel illustrates the proximity of G1218, E1219, and T1337 to R1335, a residue that makes a direct, base-specific major groove contact to the third base position of the PAM¹². Angstrom distances indicated by yellow dashed lines; non-target strand guanine bases dG2 and dG3 of the PAM are shown in blue; other DNA bases shown in orange; water molecules shown in red; images generated using

PyMOL from PDB:4UN3. **b**, Mutational analysis of six residues in SpCas9 that are implicated in PAM recognition. Clones containing one of three types of mutations at each position were tested for EGFP disruption with two sgRNAs targeted to sites harbouring NGG PAMs. For each position, we created an alanine substitution and two non-conservative mutations. S1136 and R1335 were previously reported to mediate contacts to the third guanine of the PAM¹², and D1135, G1218, E1219, and T1337 are reported in this study. EGFP disruption activities were quantified by flow cytometry; background control represented by the dashed red line; error bars represent s.e.m., $n = 3$.

th1 - Mutations in 15/17 sequences

CGTAAGGAGGCGCAGGCGGGCGGCCGCGCGCGGAGGCTGCAGGAC TGAGCGAGCAGATCGTGTGTTTGAGG Wild-type

CGTAAGGAGGCGC-----AGCAGATCGTGTGTTTGAGG -41

CGTAAGGAGGCGCGAGG-----CGAGCAGATCGTGTGTTTGAGG -34

CGTAAGGAGGCGCGAGGCG-----TGAGCAGCAGATCGTGTGTTTGAGG -29

CGTAAGGAGGCGCAGGCGGGCGGC-----GAGCAGATCGTGTGTTTGAGG -28

CGTAAGGAGGCGCGAGGCGGGCGGCCG agatc-----GAGCAGCAGATCGTGTGTTTGAGG -16 (-21,+5)

CGTAAGGAGGCGCGAGGCGGGCGGCCGCGCGCGC t-----AGCAGCAGATCGTGTGTTTGAGG -15 (-16,+1)

CGTAAGGAGGCGCAGGCGGGCGGCCGCGCGCGCGG-----GCAGCAGATCGTGTGTTTGAGG -15

CGTAAGGAGGCGCGAGGCGGGCGGCCGCGCGGC-----TGAGCAGCAGATCGTGTGTTTGAGG -14

CGTAAGGAGGCGCGAGGCG agggcgccggcg-----GCAGCAGATCGTGTGTTTGAGG -13 (-32,+19)

CGTAAGGAGGCGC agagagcgtaaaggagcgcgaggcg-----GCAGCAGATCGTGTGTTTGAGG -13 (-37,+24)

CGTAAGGAGGCGCAGGCGGGCGGCCGCGCGCGCGAGG-----CTAGCAGCAGATCGTGTGTTTGAGG -8

CGTAAGGAGGCGCAGGCGGGCGGCCGCGCGCGCGGAGGCTGCAG-ACTGAGCAGCAGATCGTGTGTTTGAGG -1 [2x]

CGTAAGGAGGCGCGAGGCGGGCGGCCGCGCGCGCGAGGCTGC agcggaCTAGCAGCAGCAGATCGTGTGTTGA +2 (-3,+5)

CGTAAGGAGGCGCAGGCGGGCGGCCGCGCGCGCGAGGCTGCAG cgaaacaaacgaacaaatcCGTGTGTTGA +2 (-17,+19)

tia11 - Mutations in 17/27 sequences

TGTCGGGAACCTCTCTCAGGGATGTTACGGAGGCCCTCATCTCTGCAAGTGTTCTCTCAGATC Wild-type

TGTCGGGtat-----GTTACGGAGGCCCTCATCTCTGCAAGTGTTCTCTCAGATC -12 (-15,+3)

TGTCGGGAACCTCTCT-----GTTACGGAGGCCCTCATCTCTGCAAGTGTTCTCTCAGATC -8 [X4]

TGTCGGGAACCTCTCTC-----TGTTACGGAGGCCCTCATCTCTGCAAGTGTTCTCTCAGATC -5

TGTCGGGAACCTCTCCA-----TGTTACGGAGGCCCTCATCTCTGCAAGTGTTCTCTCAGATC -4 [X3]

TGTCGGGAACCTCTCTC-----GGGATGTTACGGAGGCCCTCATCTCTGCAAGTGTTCTCTCAGATC -1

TGTCGGGAACCTCTCTCAT-----GATGTTACGGAGGCCCTCATCTCTGCAAGTGTTCTCTCAGATC -1 (-2,+1)

TGTCGGGAACCTCTCTCAGGAGCTTACGGAGGCCCTCATCTCTGCAAGTGTTCTCTCAGATC 0 (-1,+1)

TGTCGGGAACCTCTCTgaacCGGATGTTACGGAGGCCCTCATCTCTGCAAGTGTTCTCTCAGATC +1 (-4,+5)

TGTCGGGAACCTCTCTgttacGAGTGTTACGGAGGCCCTCATCTCTGCAAGTGTTCTCTCCTA +4 (-4,+8)

TGTCGGGAACCTCTCTCattgtacGGATGTTACGGAGGCCCTCATCTCTGCAAGTGTTCTCTC +5 (-1,+6)

TGTCGGGAACCTCTCTCattgttgGATGTTACGGAGGCCCTCATCTCTGCAAGTGTTCTCTC +5 (-1,+6)

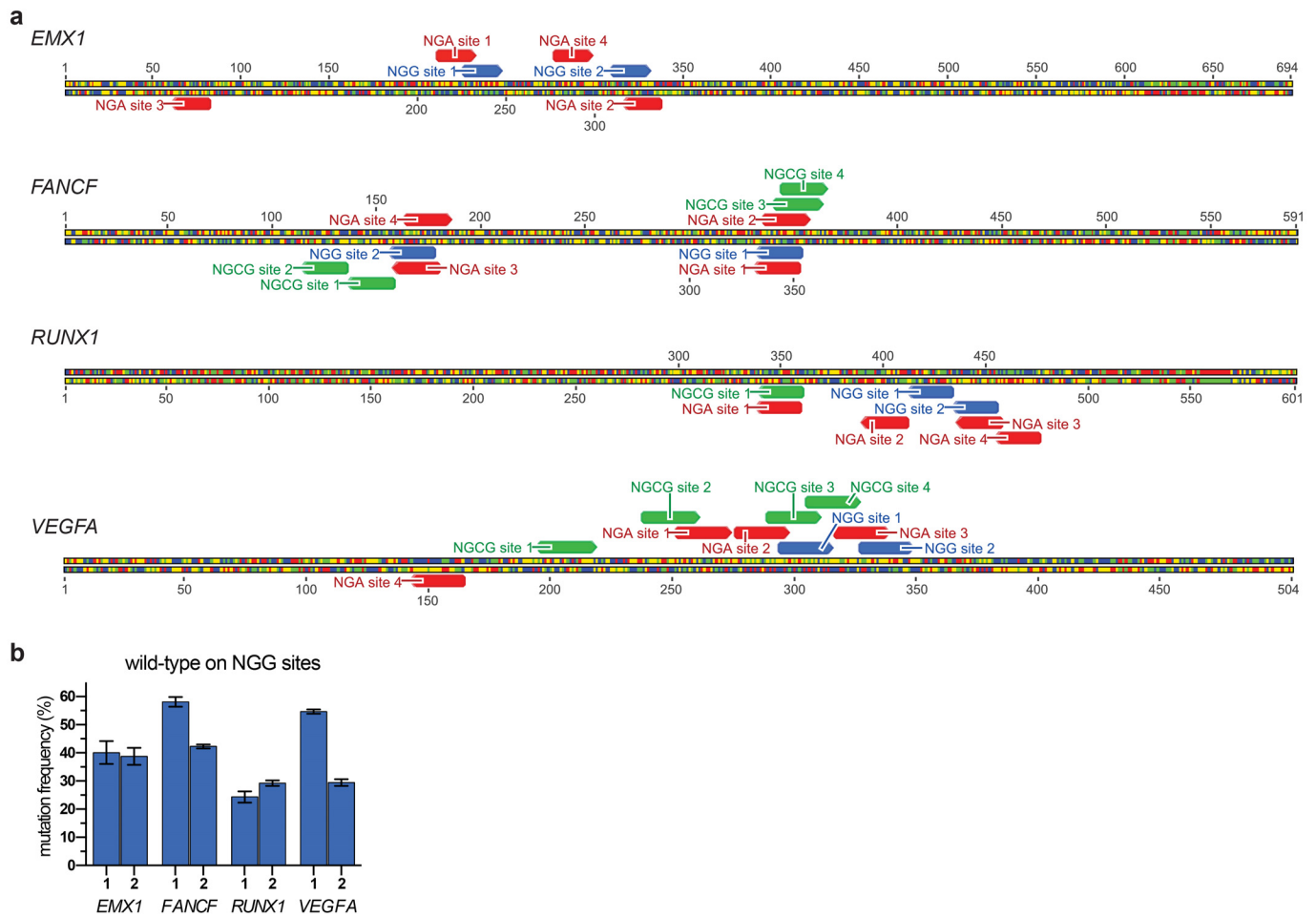
TGTCGGGAACCTCTCTCatttgaggaattattataaaataaTACCGAGGCCCT +20 (-11,+31)

fh - Mutations in 6/20 sequences

CATGGCGACCGGGGGC**GGA**ACTACTGCTCT**CCAAC****AGAG**GCGAGAATCGGGGGCGGACG Wild-type
CATGGCGACCGGGGGCGGA**ACTACTGC**-----ACCAGAGGCGAGAATCGGGGGCGGACG -6
CATGGCGACCGGGGGCGGA**ACTACTGCTCT**-----CAGAGGCGAGAATCGGGGGCGGACG -5
CATGGCGACCGGGGGCGGA**ACTACTGCTCT**-----CCAGAGGCGAGAATCGGGGGCGGACG -4
CATGGCGACCGGGGGCGGA**ACTACTGCTCg**-----CCAGAGGCGAGAATCGGGGGCGGACG -4 (-5, +1)
CATGGCGACCGGGGGCGGA**ACTACTGCTCgagagc**CCAGAGGCGAGAATCGGGGGCGGACG +3 (-3, +6)
CATGGCGACCGGGGGCGGA**AGCTACTGCTCTctactgctctctt**CCAGAGGCGAGAATCGGCG +10 (-2, +12)

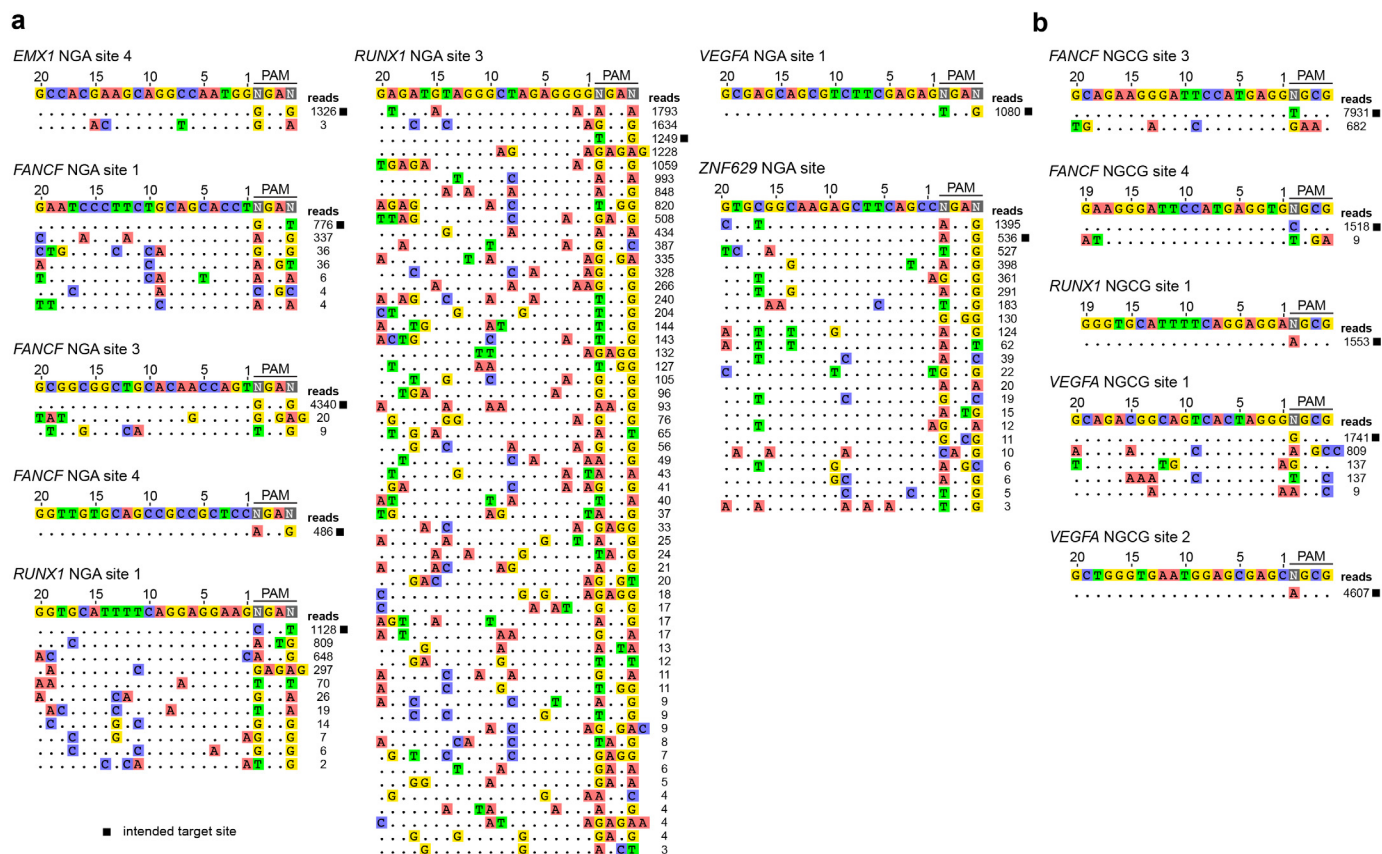
Extended Data Figure 6 | Insertion or deletion mutations induced by the VQR SpCas9 variant at endogenous zebrafish sites containing NGAG PAMs. For each target locus, the wild-type sequence is shown at the top with the protospacer highlighted in yellow (highlighted in green if present on the complementary strand) and the PAM is marked as red underlined text. Deletions are shown as red dashes highlighted in grey and insertions as lower

case letters highlighted in blue. The net change in length caused by each indel mutation is shown on the right (+, insertion; -, deletion). Note that some alterations have both insertions and deletions of sequence and in these instances the alterations are enumerated in parentheses. The number of times each mutant allele was recovered (if more than once) is shown in brackets.



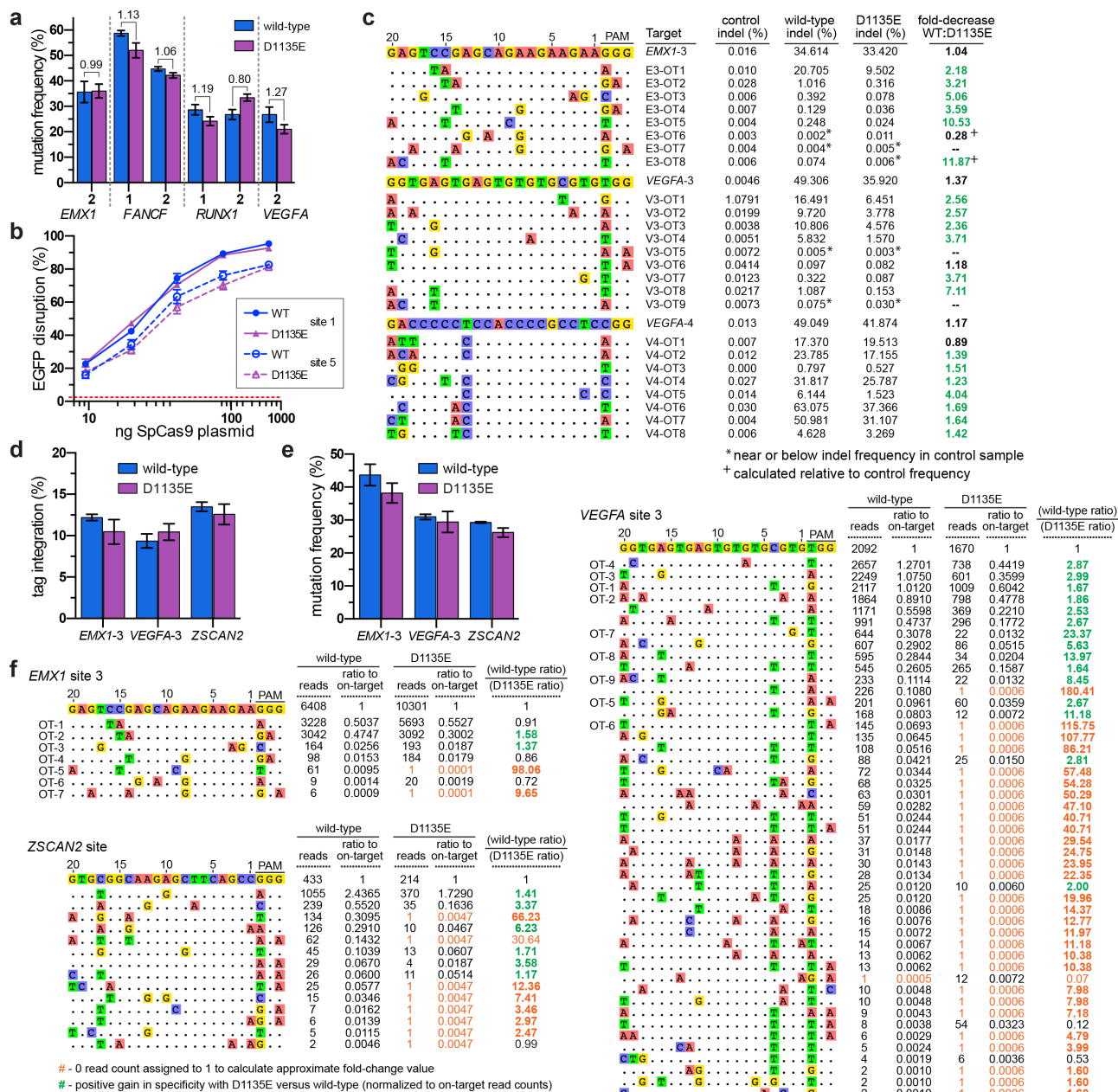
Extended Data Figure 7 | Endogenous human genes targeted by wild-type and evolved variants of SpCas9. a, Sequences targeted by wild-type, VQR, and VRER SpCas9 are shown in blue, red, and green, respectively. Sequences of sgRNAs and primers used to amplify these loci for T7E1 are provided in

Supplementary Tables 1 and 2. **b,** Mean mutagenesis frequencies detected by T7E1 for wild-type SpCas9 at eight target sites bearing NGG PAMs in the four different endogenous human genes (corresponding to the annotations in panel a). Error bars represent s.e.m., $n = 3$.



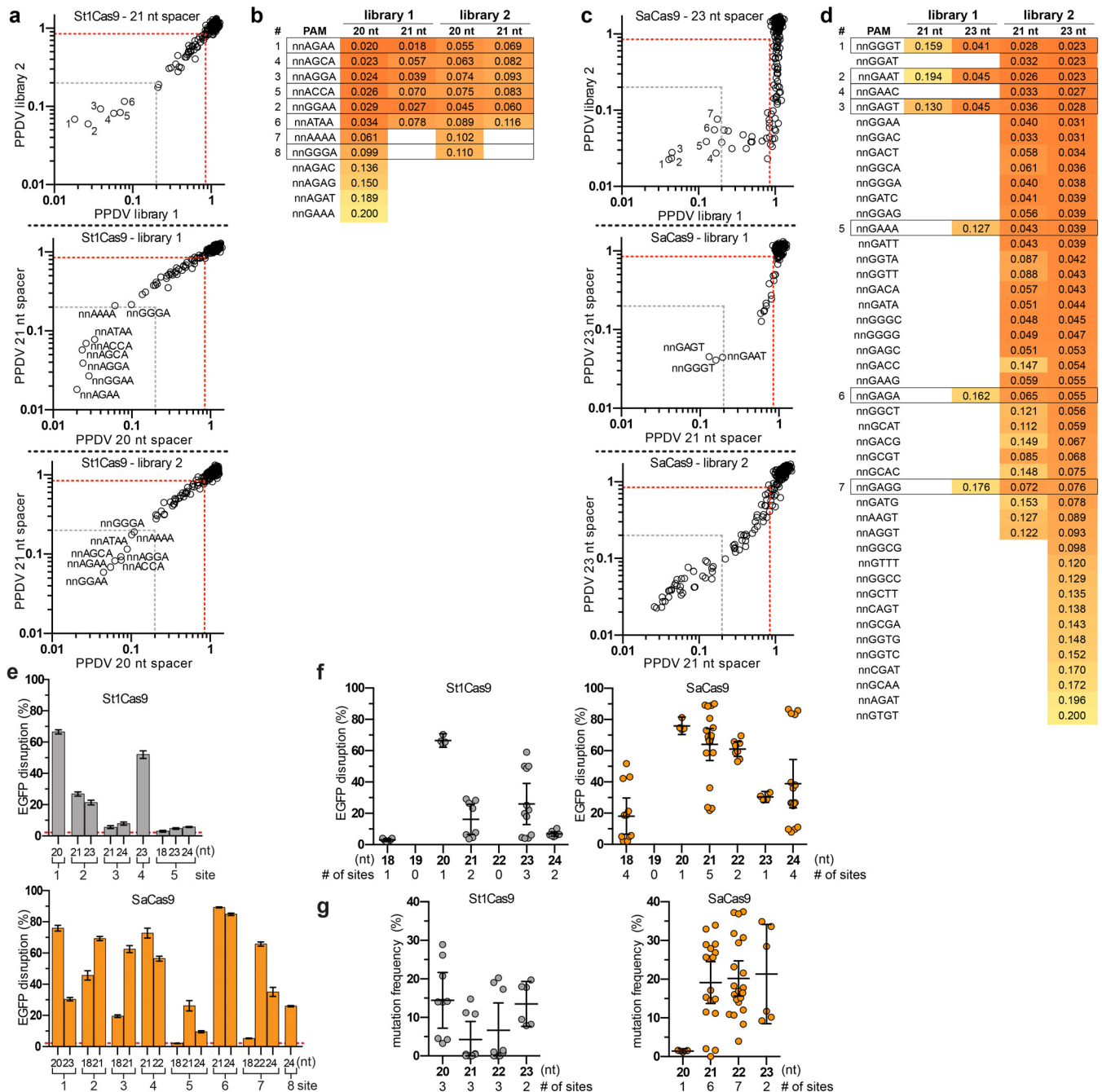
Extended Data Figure 8 | Specificity profiles of the VQR and VRER SpCas9 variants determined using GUIDE-seq⁷. The intended on-target site is marked with a black square, and mismatched positions within off-target sites are highlighted. **a**, The specificity of the VQR variant was assessed in human cells by targeting endogenous sites containing NGA PAMs: *EMX1* site 4,

FANCF site 1, *FANCF* site 3, *FANCF* site 4, *RUNX1* site 1, *RUNX1* site 3, *VEGFA* site 1, and *ZNF629*. **b**, The specificity of the VRER variant was assessed in human cells by targeting endogenous sites containing NGCG PAMs: *FANCF* site 3, *FANCF* site 4, *RUNX1* site 1, *VEGFA* site 1, and *VEGFA* site 2.



Extended Data Figure 9 | Activity differences between D1135E and wild-type SpCas9. **a**, Mutagenesis frequencies detected by T7E1 for wild-type and D1135E SpCas9 at six endogenous sites in human cells. Error bars represent s.e.m., $n = 3$; mean fold change in activity is shown. **b**, Titration of the amount of wild-type or D1135E SpCas9-encoding plasmid transfected for EGFP disruption experiments in human cells. The amount of sgRNA plasmid used for all of these experiments was fixed at 250 ng. Two sgRNAs targeting different EGFP sites were used; error bars represent s.e.m., $n = 3$. **c**, Targeted deep-sequencing of on- and off-target sites for 3 sgRNAs using wild-type and D1135E SpCas9. The on-target site is shown at the top, with off-target sites listed below highlighting mismatches to the on-target. Fold decreases in activity with D1135E relative to wild-type SpCas9 at off-target sites greater than the change in activity at the on-target site are highlighted in green; control indel levels for each amplicon are reported. **d**, Mean frequency of GUIDE-seq oligo

tag integration at the on-target sites, estimated by restriction fragment length polymorphism analysis. Error bars represent s.e.m., $n = 4$. **e**, Mean mutagenesis frequencies at the on-target sites detected by T7E1 for GUIDE-seq experiments. Error bars represent s.e.m., $n = 4$. **f**, GUIDE-seq read count comparison between wild-type SpCas9 and D1135E at 3 endogenous human cell sites. The on-target site is shown at the top and off-target sites are listed below with mismatches highlighted. In the table, a ratio of off-target activity to on-target activity is compared between wild-type and D1135E to calculate the normalized fold changes in specificity (with gains in specificity highlighted in green). For sites without detectable GUIDE-seq reads, a value of 1 has been assigned to calculate an estimated change in specificity (indicated in orange). Off-target sites analysed by deep-sequencing in panel **c** are numbered to the left of the *EMX1* site 3 and *VEGFA* site 3 off-target sites.



Extended Data Figure 10 | Additional PAMs for St1Cas9 and SaCas9 and activities based on spacer lengths in human cells. **a**, PPDV scatterplots for St1Cas9 comparing the sgRNA complementarity lengths of 20 and 21 nucleotides obtained with a randomized PAM library for spacers 1 and 2 (see also Fig. 4a). PAMs were grouped and plotted by their third/fourth/fifth/sixth positions. The red dashed line indicates PAMs that are statistically significantly depleted (see Extended Data Fig. 3c) and the grey dashed line represents fivefold depletion (PPDV of 0.2). **b**, Table of PAMs with PPDVs of less than 0.2 for St1Cas9 under each of the four conditions tested. PAM numbering shown on the left is the same as in Fig. 4a. **c**, PPDV scatterplots for SaCas9 comparing the sgRNA complementarity lengths of 21 and 23 nucleotides obtained with a randomized PAM library for spacers 1 and 2 (see also Fig. 4b). PAMs were

grouped and plotted by their third/fourth/fifth/sixth positions. The red and grey dashed lines are the same as in **a**. **d**, Table of PAMs with PPDVs of less than 0.2 for SaCas9 under each of the four conditions tested. PAM numbering shown on the left is the same as in Fig. 4b. **e**, Human cell EGFP disruption activities of St1Cas9 and SaCas9 at sites of various spacer lengths. Frequencies were quantified by flow cytometry; error bars represent s.e.m., $n = 3$ or 4 ; mean level of background EGFP loss represented by the dashed red line. **f**, Activity for all replicates of data shown in **e** plotted against spacer length. $n = 3$ or 4 ; bars illustrate mean and 95% confidence interval; number of sites per spacer length indicated. **g**, Activity for all replicates shown in Fig. 4d, **e**, plotted against spacer length. $n = 3$ or 4 ; bars illustrate mean and 95% confidence interval; number of sites per spacer length indicated.

Single-cell chromatin accessibility reveals principles of regulatory variation

Jason D. Buenrostro^{1,2}, Beijing Wu^{1*}, Ulrike M. Litzénburger^{2*}, Dave Ruff³, Michael L. Gonzales³, Michael P. Snyder¹, Howard Y. Chang² & William J. Greenleaf^{1,4}

Cell-to-cell variation is a universal feature of life that affects a wide range of biological phenomena, from developmental plasticity^{1,2} to tumour heterogeneity³. Although recent advances have improved our ability to document cellular phenotypic variation^{4–8}, the fundamental mechanisms that generate variability from identical DNA sequences remain elusive. Here we reveal the landscape and principles of mammalian DNA regulatory variation by developing a robust method for mapping the accessible genome of individual cells by assay for transposase-accessible chromatin using sequencing (ATAC-seq)⁹ integrated into a programmable microfluidics platform. Single-cell ATAC-seq (scATAC-seq) maps from hundreds of single cells in aggregate closely resemble accessibility profiles from tens of millions of cells and provide insights into cell-to-cell variation. Accessibility variance is systematically associated with specific *trans*-factors and *cis*-elements, and we discover combinations of *trans*-factors associated with either induction or suppression of cell-to-cell variability. We further identify sets of *trans*-factors associated with cell-type-specific accessibility variance across eight cell types. Targeted perturbations of cell cycle or transcription factor signalling evoke stimulus-specific changes in this observed variability. The pattern of accessibility variation in *cis* across the genome recapitulates chromosome compartments¹⁰ *de novo*, linking single-cell accessibility variation to three-dimensional genome organization. Single-cell analysis of DNA accessibility provides new insight into cellular variation of the ‘regulome’.

Heterogeneity within cellular populations has been evident since the first microscopic observations of individual cells. Recent proliferation of powerful methods for interrogating single cells^{4–8} has allowed detailed characterization of this molecular variation, and provided deep insight into characteristics underlying developmental plasticity^{1,2}, cancer heterogeneity³, and drug resistance¹¹. In parallel, genome-wide mapping of regulatory elements in large ensembles of cells have unveiled substantial variation in chromatin structure across cell types, particularly at distal regulatory regions¹². In particular, methods for probing genome-wide DNA accessibility have proven extremely effective in identifying regulatory elements across a variety of cell types¹³ and quantifying changes that lead to both activation or repression of gene expression. Given this broad diversity of activity within regulatory elements when comparing phenotypically distinct cell populations, it is reasonable to hypothesize that heterogeneity at the single-cell level extends to accessibility variability within cell types at regulatory elements. However, the lack of methods to probe DNA accessibility within individual cells has prevented quantitative dissection of this hypothesized regulatory variation.

We have developed a single-cell assay for transposase-accessible chromatin (scATAC-seq). ATAC-seq is an ensemble measure of open chromatin that uses the prokaryotic Tn5 transposase^{14,15} to tag regulatory regions by inserting sequencing adapters into accessible regions of the genome. In scATAC-seq, individual cells are

captured and assayed using a programmable microfluidics platform (Fluidigm) with methods optimized for this task (Fig. 1a, Extended Data Fig. 1 and Supplementary Discussion). After transposition and PCR on the integrated fluidics circuit (IFC), libraries were collected and PCR amplified with cell-identifying barcoded primers. Single-cell libraries were then pooled and sequenced on a high-throughput sequencing instrument. Using single-cell ATAC-seq, we generated DNA accessibility maps from 254 individual GM12878 lymphoblastoid cells. Aggregate profiles of scATAC-seq data closely reproduce ensemble measures of accessibility profiled by DNase-seq and ATAC-seq generated from $\sim 10^7$ or $\sim 10^4$ cells, respectively (Fig. 1b, c and Extended Data Fig. 2a). Data from single cells recapitulate several characteristics of bulk ATAC-seq data, including fragment-size periodicity corresponding to integer multiples of nucleosomes, and a strong enrichment of fragments within regions of accessible chromatin (Extended Data Fig. 2b, c). Microfluidic chambers generating low library diversity or poor measures of accessibility, which correlate with empty chambers or dead cells, were excluded from further analysis (Fig. 1d and Extended Data Fig. 2d–l). Chambers passing filter yielded an average of 7.3×10^4 fragments mapping to the nuclear genome. We further validated the approach by measuring chromatin accessibility from a total of 1,632 IFC chambers representing three tier 1 ENCODE cell lines¹⁶ (H1 human embryonic stem cells (ES cells), K562 chronic myelogenous leukaemia and GM12878 lymphoblastoid cells), as well as from V6.5 mouse ES cells, EML¹ cells (mouse haematopoietic progenitors), TF-1 cells (human erythroblast), HL-60 cells (human promyeloblasts) and BJ fibroblasts (human foreskin fibroblasts).

Because regulatory elements are generally present at two copies in a diploid genome, we observe a near digital (0 or 1) measurement of accessibility at individual elements within individual cells (Extended Data Fig. 3a). For example, within a typical single cell we estimate a total of 9.4% of promoters are represented in a typical scATAC-seq library (Extended Data Fig. 3b–d). The sparse nature of scATAC-seq data makes analysis of cellular variation at individual regulatory elements impractical. We therefore developed an analysis infrastructure to measure regulatory variation using changes of accessibility across sets of genomic features (Fig. 2a, b). To quantify this variation we first choose a set of open chromatin peaks, identified using the aggregate accessibility track, which share a common characteristic (such as transcription factor binding motif, ChIP-seq peaks or cell cycle replication timing domains). We then calculate the observed fragments in these regions minus the expected fragments, downsampled from the aggregate profile, within individual cells. To correct for bias, we divide this by the root mean square of fragments expected from a background signal constructed to estimate technical and sampling error within single-cell data sets (Methods and Extended Data Fig. 4). Hereafter, we refer to this metric as ‘deviation’. Finally, for any set of features, we also calculate an overall ‘variability’ score across

¹Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA. ²Program in Epithelial Biology and the Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California 94305, USA. ³Fluidigm Corporation, South San Francisco, California 94080, USA. ⁴Department of Applied Physics, Stanford University, Stanford, California 94025, USA.

*These authors contributed equally to this work.

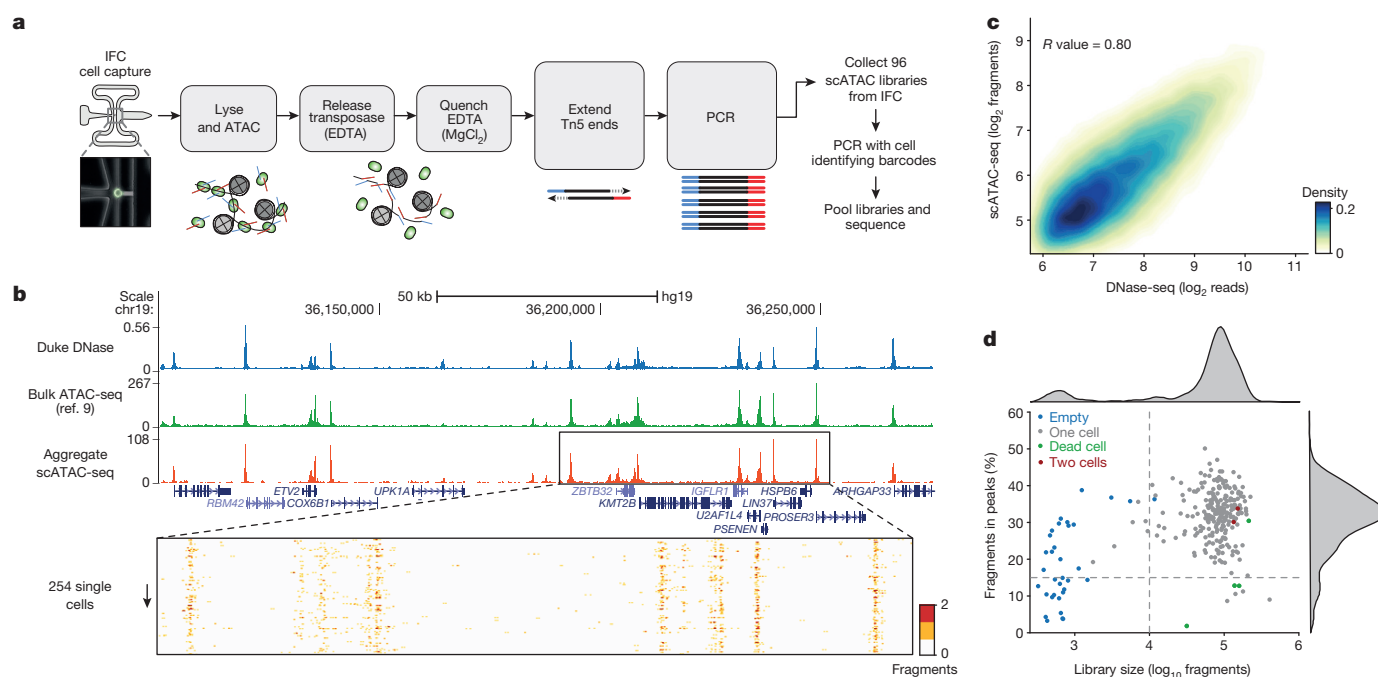


Figure 1 | Single-cell ATAC-seq provides an accurate measure of chromatin accessibility genome-wide. **a**, Workflow for measuring single epigenomes using scATAC-seq on a microfluidic device (Fluidigm). **b**, Aggregate single-cell accessibility profiles closely recapitulate profiles of DNase-seq and ATAC-seq in GM12878 cells. **c**, Genome-wide accessibility patterns observed by

scATAC-seq are correlated with DNase-seq data ($R = 0.80$). **d**, Library size versus percentage of fragments in open chromatin peaks (filtered as described in Methods) within K562 cells ($n = 288$). Dotted lines (15% and 10,000) represent cutoffs used for downstream analysis.

all cells (Fig. 2b), a metric of excess variance over the background signal.

We first focused our analysis on K562 myeloid leukaemia cells, a cell type with extensive epigenomic data sets^{17,18}. To comprehensively characterize variability associated with *trans*-factors within individual K562 cells, we computed variability across all available ENCODE ChIP-seq, transcription factor motifs and regions that differed in replication timing (as determined from Repli-Seq data sets¹⁹) (Fig. 2c, d). We found measures of cell-to-cell variability were highly reproducible across biological replicates (Extended Data Fig. 5). As expected from proliferating cells, we find increased variability within different replication timing domains, representing variable ATAC-seq signal associated with changes in DNA content across the cell cycle. In addition, we discover a set of *trans*-factors associated with high variability. These factors include sequence-specific transcription factors, such as GATA1/2, JUN and STAT2, and chromatin effectors, such as BRG1 (also known as SMARCA4) and P300 (also known as EP300). Immunostaining followed by microscopy or flow cytometry (Fig. 2e and Extended Data Fig. 6a–d) confirmed heterogeneous expression of GATA1 and GATA2. Principal component (PC) analysis of single-cell deviations across all *trans*-factors show seven significant PCs, with PC 5 describing changes in DNA abundance throughout the cell cycle. This analysis suggests that high-variance *trans*-factors are variable independent of the cell cycle (Fig. 2f and Extended Data Fig. 6e–g). The remaining PCs show contributions from several transcription factors, suggesting that variance across sets of *trans*-factors represent distinct regulatory states in individual cells.

We hypothesized that variation associated with different *trans*-factors can synergize, either through cooperative or competitive binding, to induce or suppress site-to-site variability in chromatin accessibility. For example, the most variant factors in K562 cells, GATA1 and GATA2, display expression heterogeneity and also bind an identical consensus sequence GATA, suggesting these factors may compete for access to DNA sequences. In support of this hypothesis, we find

regulatory elements with both GATA1 and GATA2 ChIP-seq signals show increased variability in accessibility, whereas sites with only GATA1 or GATA2 show substantially less variability (Fig. 2g and Extended Data Fig. 6h). In contrast, we find no substantial change in variability of GATA1 binding sites that co-occur with JUN or CEBPB (Extended Data Fig. 6i). We also find peaks unique to GATA1 binding are significantly more accessible than peaks unique to GATA2 (Extended Data Fig. 6k–l) supporting the hypothesis that GATA1, an activator of accessibility, competes with GATA2 to induce single-cell variability. Extending this analysis to all transcription factor ChIP-seq data sets revealed a *trans*-factor synergy landscape for accessibility variation (Fig. 2g and Extended Data Fig. 6j). For example, chromatin accessibility variance associated with GATA2 binding is significantly enhanced when the same region could also be bound by GATA1, TAL1 or P300. In contrast, CTCF, SUZ12, and ZNF143 appear to act as general suppressors of accessibility variance, unless associated with proximal binding of ZNF143 or SMC3, the latter a cohesin subunit involved in chromosome looping^{18,20}. Thus, single cell accessibility profiles nominate distinct *trans*-factors that, in combination, induce or suppress cell-to-cell regulatory variation.

To validate our ability to detect changes in accessibility variance, we used chemical inhibitors to modulate potential sources of cell-cell variability. Inhibition of cyclin-dependent kinases 4 and 6 (CDK4/6), essential components of the cell cycle, caused a marked reduction of variability within peaks associated with DNA replication timing domains (Repli-Seq) (Fig. 3a). The addition of inhibitors of JUN or BCR-ABL kinases (JNKi and imatinib, respectively) increased G1/S-associated variability suggesting an increase in the subpopulation of G1/S cells, which was validated with flow cytometry (Extended Data Fig. 7). JUN variability was significantly gained in response to JNKi but not imatinib treatment, suggesting that high-variance *trans*-factors can also be specifically and pharmacologically modulated. Tumour necrosis factor (TNF) treatment of GM12878 cells specifically modulated accessibility variability at NF- κ B sites (Fig. 3b), consistent with the known stochastic and oscillatory property of nuclear shuttling in

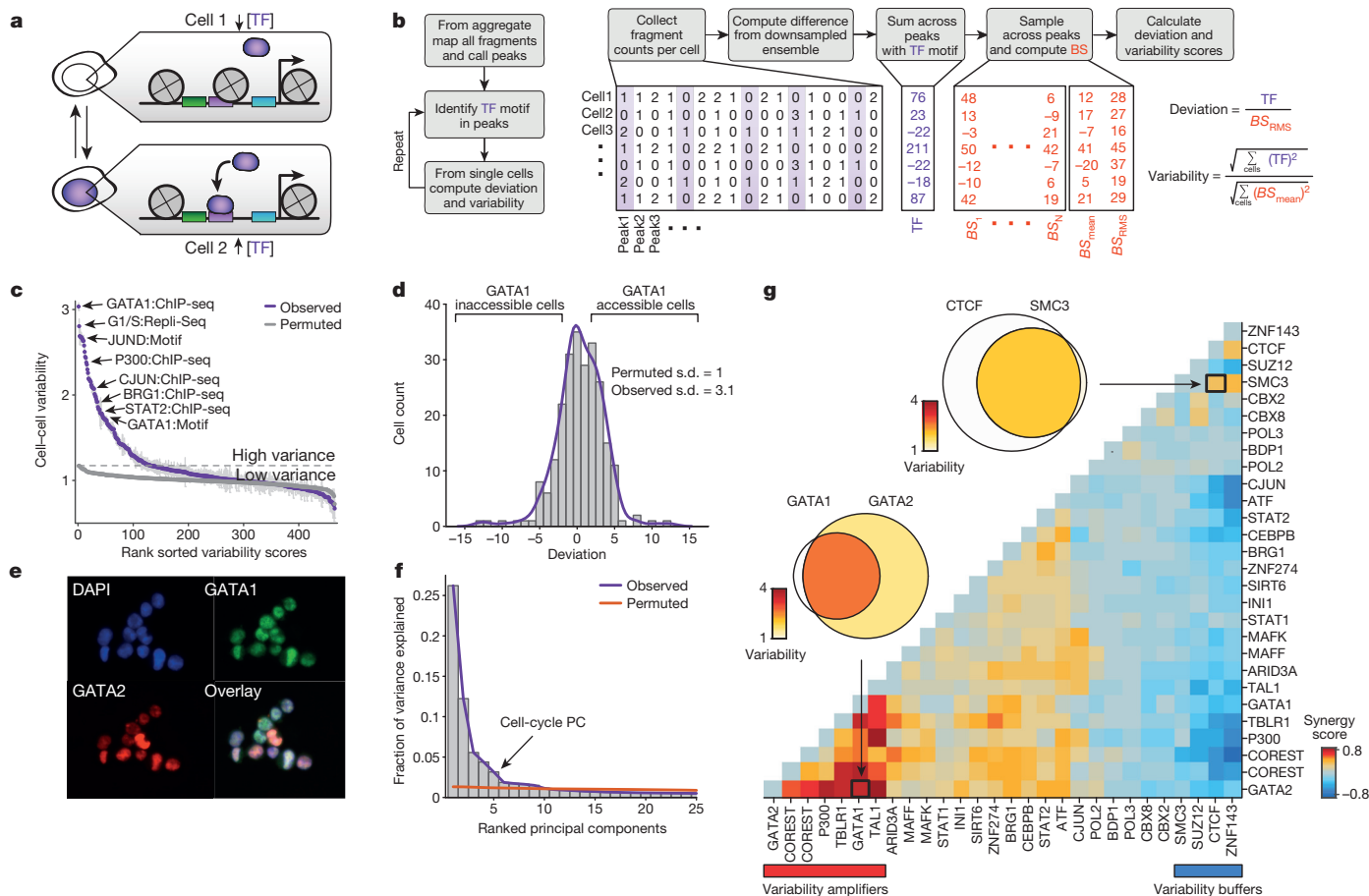


Figure 2 | Trans-factors are associated with single-cell epigenomic variability. **a**, Schematic showing two cellular states (transcription factor high and transcription factor low) leading to differential chromatin accessibility. TF, transcription factor. **b**, Analysis infrastructure, which uses a calculated background signal (BS; see Supplementary Methods, section 3.2) to calculate transcription factor deviations and variability from scATAC-seq data. The transcription factor value is calculated by subtracting the number of expected fragments from the observed fragments per cell (see Supplementary Methods, section 3.1). **c**, Observed cell-to-cell variability within sets of genomic features associated with ChIP-seq peaks, transcription factor motifs, and replication timing (error estimates shown in grey, see Methods for details). Variability

this system²¹. Together, these results show that variability can be experimentally modulated and further demonstrates that variability is not solely dependent on the cell cycle.

We observe that *trans*-factors associated with high variability are generally cell-type specific. Hierarchical bi-clustering of single-cell deviations generated from three cell lines reveals cell-type specific sets of transcription factor motifs associated with high variability (Fig. 3c). This analysis also shows cells from different biological replicates cluster with their cell type of origin (with a single exception), suggesting scATAC-seq can also be used to deconvolve heterogeneous cellular mixtures. Systematic analysis of all assayed cell types identified high-variance *trans*-factor motifs that are generally unique to specific cell types (Fig. 3d and Extended Data Fig. 8a). For example, regions associated with GATA transcription factors are most variant in K562 cells, whereas regions associated with master pluripotency transcription factors Nanog and Sox2 are most variant in mouse ES cells, consistent with previous observations of expression variation of these factors^{22,23}. We also find high variability of GATA1 and PU.1 (SPI1) binding accessibility in EML cells, a cell type previously shown to have >200-fold GATA1 and >15-fold PU.1 expression differences within clonal cellular subpopulations¹. The complete set of identified high-variance *trans*-factors contains a number of transcription factors pre-

measured from permuted background (see Methods) is shown in grey dots.

d, Distribution of normalized deviations from expected accessibility signal for GATA1 sites in individual cells, histogram of cells shown in grey, density profile shown in purple (see Methods). **e**, Immunostaining of GATA1 (green) and GATA2 (red) shows protein expression in K562 cells. **f**, Principal components ranked by fraction of variance explained from observed deviation data (purple) and permuted data (orange). Bar plot of observed data shown in grey. **g**, Calculated changes in associated variability of factors when present together versus independently, depicting a context-specific *trans*-factor variability landscape (see Methods). Venn-diagrams show variability associated with GATA1 and/or GATA2 and CTCF and/or SMC3 (co-)occurring ChIP-seq sites.

viously reported to dynamically localize into the nucleus, including NF- κ B, JUN and ETS/ERG^{21,24,25}, suggesting that temporal fluctuations in transcription factor concentration may be driving observed chromatin accessibility heterogeneity. Finally, we find BJ fibroblasts and HL-60 cells exhibit less variance among this set of annotated *trans*-factor motifs, suggesting differences in the global levels of *trans*-factor variability across cell lines. Specific chromatin states and histone modifications²⁶ are also sometimes associated with accessibility variation in single cells (Extended Data Fig. 8b, c). Overall these findings suggest that *trans*-factors promote cell-type specific chromatin accessibility variation genome-wide.

Patterns of variation in accessibility along the linear genome in individual cells reveal an unexpected connection to higher-order chromosome folding. We calculated single-cell deviations within sliding windows across the genome, each encompassing a fixed number of peaks ($n = 25$) (Fig. 4a). We determined which windows co-varied within individual cells by calculating the co-correlation of each window across all others within the same chromosome within individual cells (Extended Data Fig. 9a, b). We further enhanced this co-correlation matrix using a secondary correlation analysis using methods similar to those used in chromosome conformation studies¹⁰ (Methods). The resulting matrix, which identifies pairs of positions in the genome

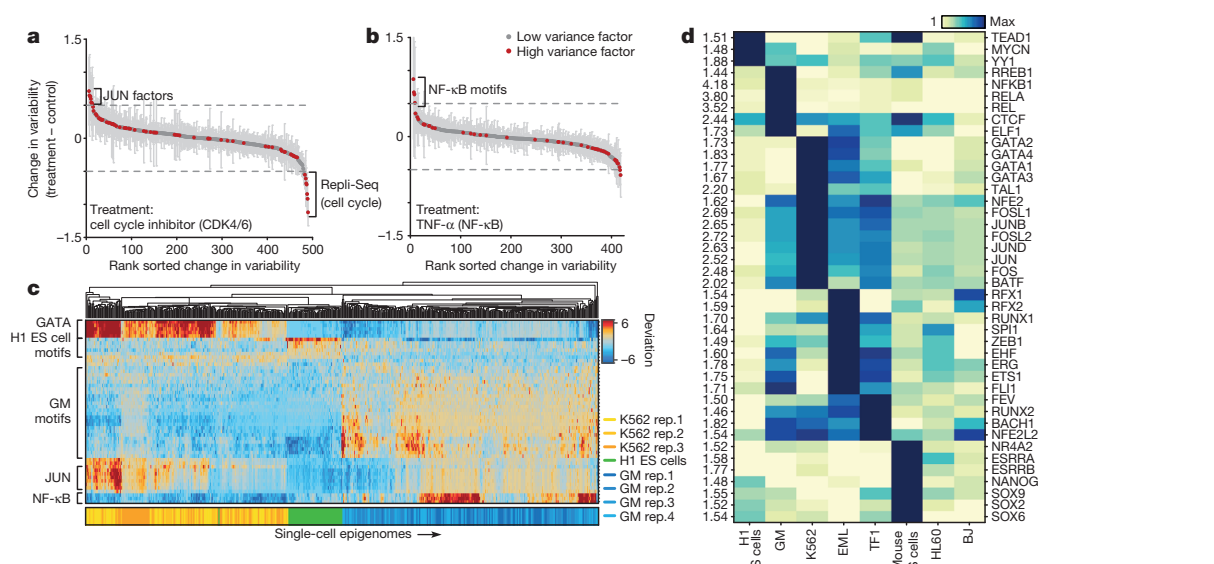


Figure 3 | Cell-type-specific epigenomic variability. **a, b,** Change of cellular variability due to chemical perturbations using CDK4/6 cell-cycle inhibitor (K562) (**a**) or TNF- α stimulation (GM12878) (**b**). Error bars (shown in grey) represent one standard deviation of bootstrapped cells across the two conditions. **c,** Heat map of deviations from expected accessibility signal across

trans-factors (rows) and of single cells (columns) from 3 cell types. Bottom colour map represents assignment classification from hierarchical clustering. **d,** Variability associated with *trans*-factor motifs across 7 cell types. Each row is normalized to the maximum variability for that motif across cell types (left).

where accessibility co-varies within individual cells, yields megabase-scale correlation domains highly concordant with previously observed chromosome compartments²⁷ (Fig. 4b–d and Extended Data Fig. 9c–i) ($R = 0.61$ for chromosome 1). These data provide independent biological validation of large-scale compartmentalization of higher-order chromatin structure^{10,27}. Moreover, these results suggest that higher-order chromatin interactions may drive regulatory variability in *cis* (elements that are proximal together tend to be accessible together). Thus, ensemble chromosome conformation data may arise in part from the statistical properties of single cell variation in co-regulated accessibility, a hypothesis also supported by single-cell fluorescent

in situ hybridization (FISH) measurements of interactions between DNA loci²⁸.

Using scATAC-seq, we dissected single-cell epigenomic heterogeneity and linked *cis*- and *trans*-effectors to variability in accessibility profiles within individual epigenomes. We identify *trans*-factors associated with increased accessibility variance, which we call high-variance *trans*-factors. Additionally, other *trans*-factors such as CTCF appear to buffer variability, perhaps by providing a stable anchor of chromatin accessibility or insulator function that dampens potential fluctuations. Conversely, co-occurrence with other factors such as P300 appears to amplify variability, perhaps due to synergistic interactions.

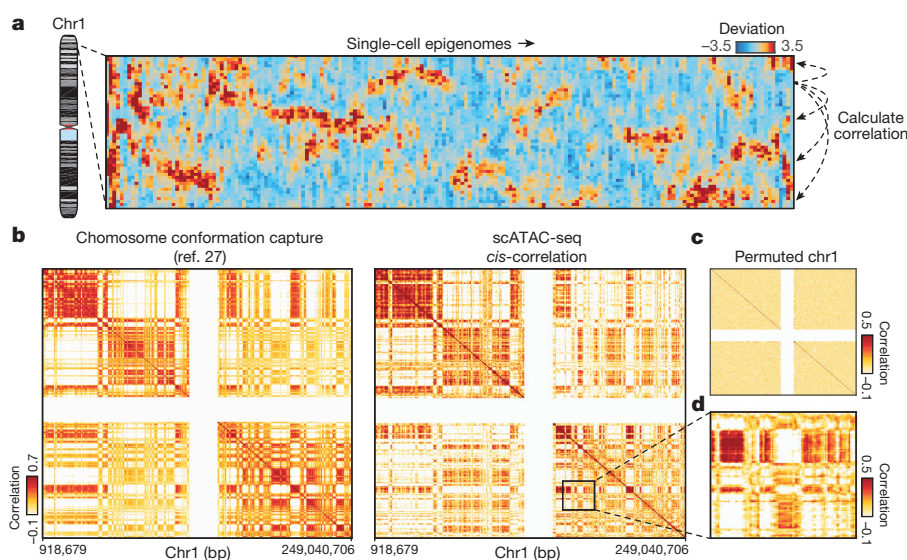


Figure 4 | Structured *cis*-variability across single epigenomes. **a,** Per-cell deviations of expected fragments across a region within chromosome 1 (see Methods). For display, only large deviation cells are shown ($n = 186$ cells). **b,** Pearson correlation coefficient representing chromosome compartment signal (see Methods) of interaction frequency from a chromatin conformation capture assay (left, analysis carried out of data from ref. 27) or doubly correlated

normalized deviations of scATAC-seq (right) from chromosome 1 (see Methods). Data in white represents masked regions due to highly repetitive regions. **c,** Permuted *cis*-correlation map for chromosome 1 (analysed identically to **b**). **d,** Box highlights a representative region depicting long-range covariability.

Lineage-specific master regulators are associated with cell-type specific single-cell epigenomic variability across several cell types, suggesting that control of single-cell variance is a fundamental characteristic of different biological states. Finally, variation of chromatin accessibility in *cis* is highly correlated with previously reported chromosome compartments, opening the intriguing possibility that this component of epigenomic noise has its roots in higher-order chromatin organization. Together these data provide a new hypothesis of regulatory mechanisms that give rise to single-cell heterogeneity.

We envision that future studies will enhance the utility of scATAC-seq by further improving the recovery of DNA fragments, increasing throughput, and refining methods of data analysis (Supplementary Discussion). Improvements to throughput and new statistical tools will enable single-cells to be partitioned by cell-state and analysed in aggregate to find the individual peaks that drive variability (Extended Data Fig. 10). In addition, we anticipate scATAC-seq may be paired with existing approaches in microscopy and single-cell RNA-seq to provide opportunities for systems analysis of individual cells. Such an approach will link regulatory variation to details of phenotypic variation, providing new insights into the molecular underpinnings of cellular heterogeneity. We believe scATAC-seq will also enable the interrogation of the epigenomic landscape of small or rare biological samples allowing for detailed, and potentially *de novo*, reconstruction of cellular differentiation or disease at the fundamental unit of investigation—the single cell.

Received 12 January; accepted 26 May 2015.

Published online 17 June 2015.

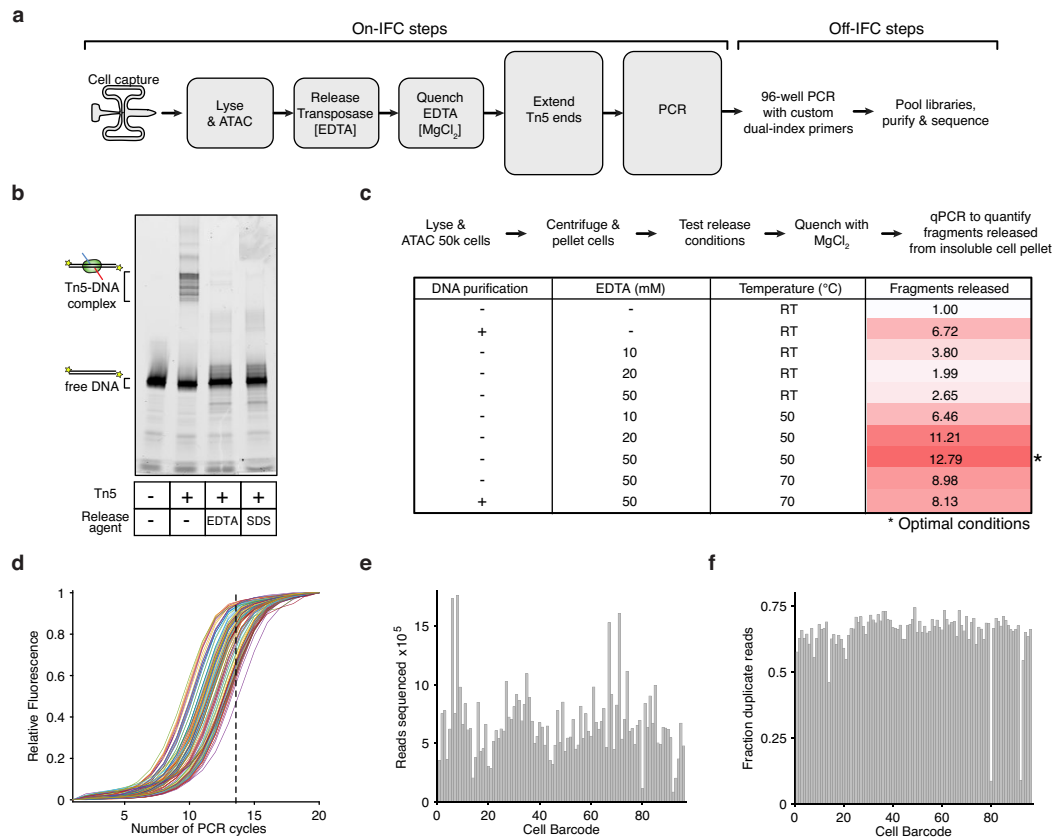
- Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. & Huang, S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**, 544–547 (2008).
- Imayoshi, I. *et al.* Oscillatory control of factors determining multipotency and fate in mouse neural progenitors. *Science* **342**, 1203–1208 (2013).
- Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
- Bendall, S. C. *et al.* Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687–696 (2011).
- Raj, A., Rifkin, S. A., Andersen, E. & van Oudenaarden, A. Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913–918 (2010).
- Jaitin, D. A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).
- Smallwood, S. A. *et al.* Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nature Methods* **11**, 817–820 (2014).
- Zong, C., Lu, S., Chapman, A. R. & Xie, X. S. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338**, 1622–1626 (2012).
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods* **10**, 1213–1218 (2013).
- Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
- Michor, F. *et al.* Dynamics of chronic myeloid leukaemia. *Nature* **435**, 1267–1270 (2005).
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
- Goryshin, I. Y. & Reznikoff, W. S. Tn5 *in vitro* transposition. *J. Biol. Chem.* **273**, 7367–7374 (1998).
- Adey, A. *et al.* Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density *in vitro* transposition. *Genome Biol.* **11**, R119 (2010).
- ENCODE Project Consortium. User's guide to the Encyclopedia of DNA Elements (ENCODE). *PLoS Biol.* **9**, e1001046 (2011).
- Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91–100 (2012).
- Xie, D. *et al.* Dynamic *trans*-acting factor colocalization in human cells. *Cell* **155**, 713–724 (2013).
- Hansen, R. S. *et al.* Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl Acad. Sci. USA* **107**, 139–144 (2010).
- Pareilho, V. *et al.* Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132**, 422–433 (2008).
- Tay, S. *et al.* Single-cell NF- κ B dynamics reveal digital activation and analogue information processing. *Nature* **466**, 267–271 (2010).
- Grün, D., Kester, L. & van Oudenaarden, A. Validation of noise models for single-cell transcriptomics. *Nature Methods* **11**, 637–640 (2014).
- Singer, Z. S. *et al.* Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol. Cell* **55**, 319–331 (2014).
- Cai, L., Dalal, C. K. & Elowitz, M. B. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455**, 485–490 (2008).
- Levine, J. H., Lin, Y. & Elowitz, M. B. Functional roles of pulsing in genetic circuits. *Science* **342**, 1193–1200 (2013).
- Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
- Kalhor, R., Tjong, H., Jayatilaka, N., Alber, F. & Chen, L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nature Biotechnol.* **30**, 90–98 (2012).
- Giorgetti, L. *et al.* Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* **157**, 950–963 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by National Institutes of Health (NIH) P50HG007735 (to H.Y.C. and W.J.G.), UH2 AR067676 and Lifespan Extension Foundation (H.Y.C.), U19AI057266 (to W.J.G.) and the Rita Allen Foundation (to W.J.G.) and the Baxter Foundation Faculty Scholar Grant (to W.J.G.); H.Y.C. is an Early Career Scientist of the Howard Hughes Medical Institute. J.D.B. acknowledges support from the National Science Foundation Graduate Research Fellowships and NIH training grant T32HG000044 for support. M.P.S. acknowledges the NIH and the National Human Genome Research Institute (NHGRI) for funding through 5U54HG00455805. We thank members of Greenleaf and Chang laboratories, as well as the Fluidigm team, including L. Xi for discussions. We acknowledge the S. Kim laboratory for assistance with FACS sorting and the C. Bustamante laboratory for help with sequencing. We also thank R. Nichols, C. Mazumdar, V. Sebastiano and V. Risca for cells.

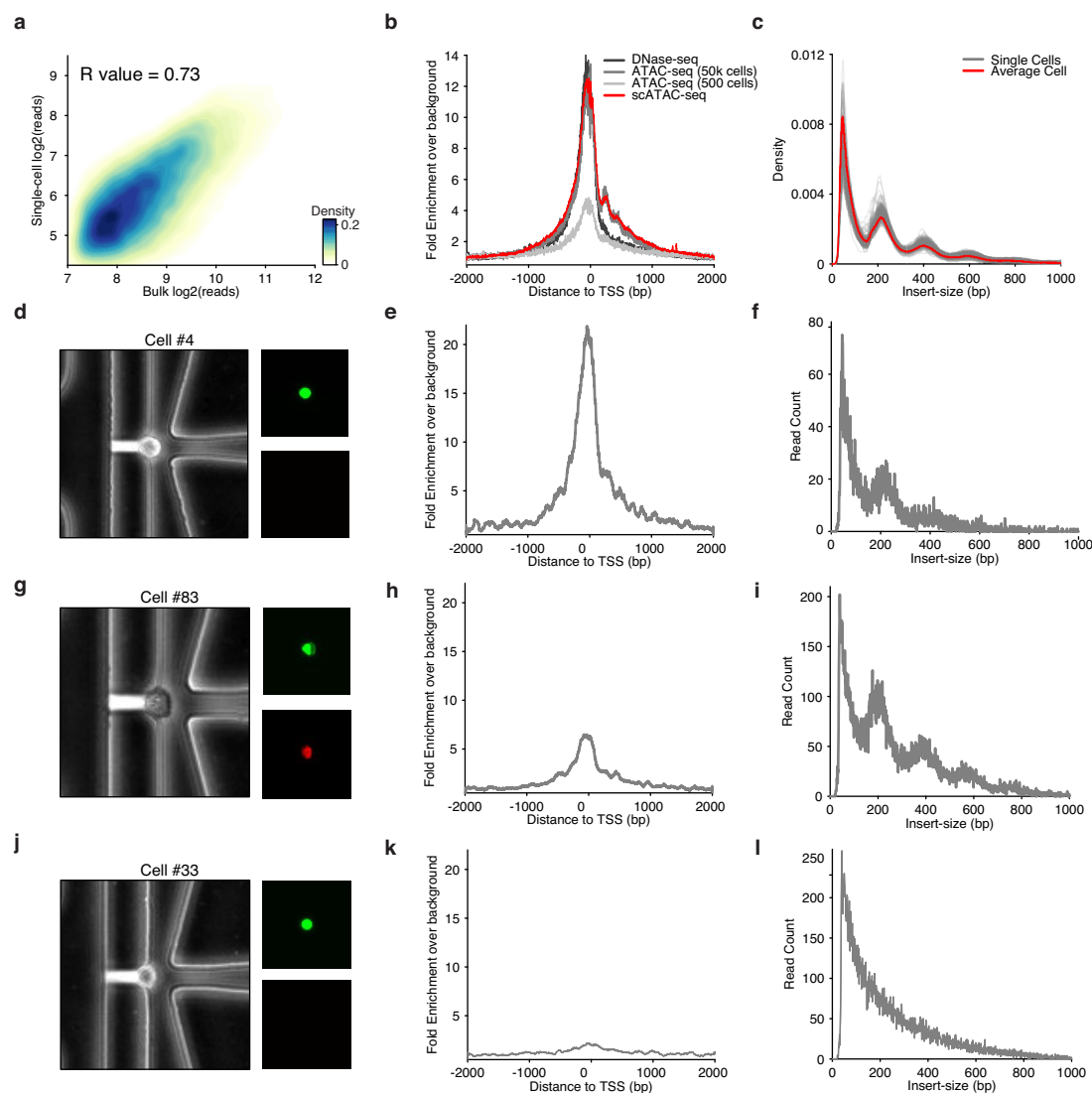
Author Contributions J.D.B., H.Y.C. and W.J.G. conceived of the method. J.D.B., B.W., M.G. and D.R. developed the Fluidigm C1 microfluidic protocols. B.W. performed all scATAC-seq experiments with supervision from J.D.B. U.M.L. conducted the flow analysis, immunostains and drug treatments. J.D.B. developed and implemented the analysis infrastructure with input from W.J.G. All authors interpreted the data and wrote the manuscript. W.J.G. and H.Y.C. supervised all aspects of this work.

Author Information All data has been deposited in GEO under the accession number GSE65360. Fluidigm C1 scripts for performing scATAC-seq are available at <https://www.fluidigm.com/c1openapp/scripthub/script/2015-06/single-cell-chromatin-accessib-1433443631246-1>. Reprints and permissions information is available at www.nature.com/reprints. Readers are welcome to comment on the online version of the paper. The authors declare competing financial interests: details are available in the online version of the paper. Correspondence and requests for materials should be addressed to W.J.G. (wjg@stanford.edu) or H.Y.C. (howchang@stanford.edu).



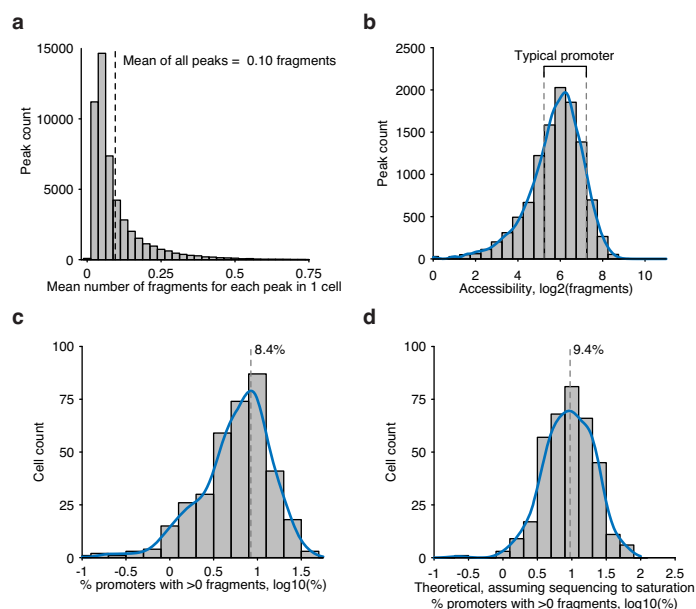
Extended Data Figure 1 | Methods development for assaying single epigenomes. **a**, scATAC-seq workflow for steps performed both on and off the integrated fluidic chip (IFC). **b**, **c**, The development of an efficient Tn5 release protocol designed to permit downstream enzymatic reactions without DNA purification. **b**, An *in vitro* electrophoretic mobility gel shift assay using a fluorescently labelled PCR product (lane 1), showing a stable Tn5-DNA complex (lane 2) dissociated with 50 mM EDTA (lane 3) or 0.1% SDS (lane 4). **c**, Workflow and associated table of conditions used to optimize release

protocol, showing conditions that markedly improve fragment yield over no release conditions or purifying DNA. Fragments released represents the fold gain in library diversity, as measured by quantitative PCR (qPCR). **d**, qPCR fluorescence traces of 96 libraries generated using scATAC-seq. For all subsequent libraries we used a total of 14 PCR cycles (dotted line). **e**, **f**, A bar plot of per-cell library sequencing depth (**e**) and fraction of duplicate reads (**f**), showing each library was sequenced to varying depths to a similar fraction of duplicate reads.



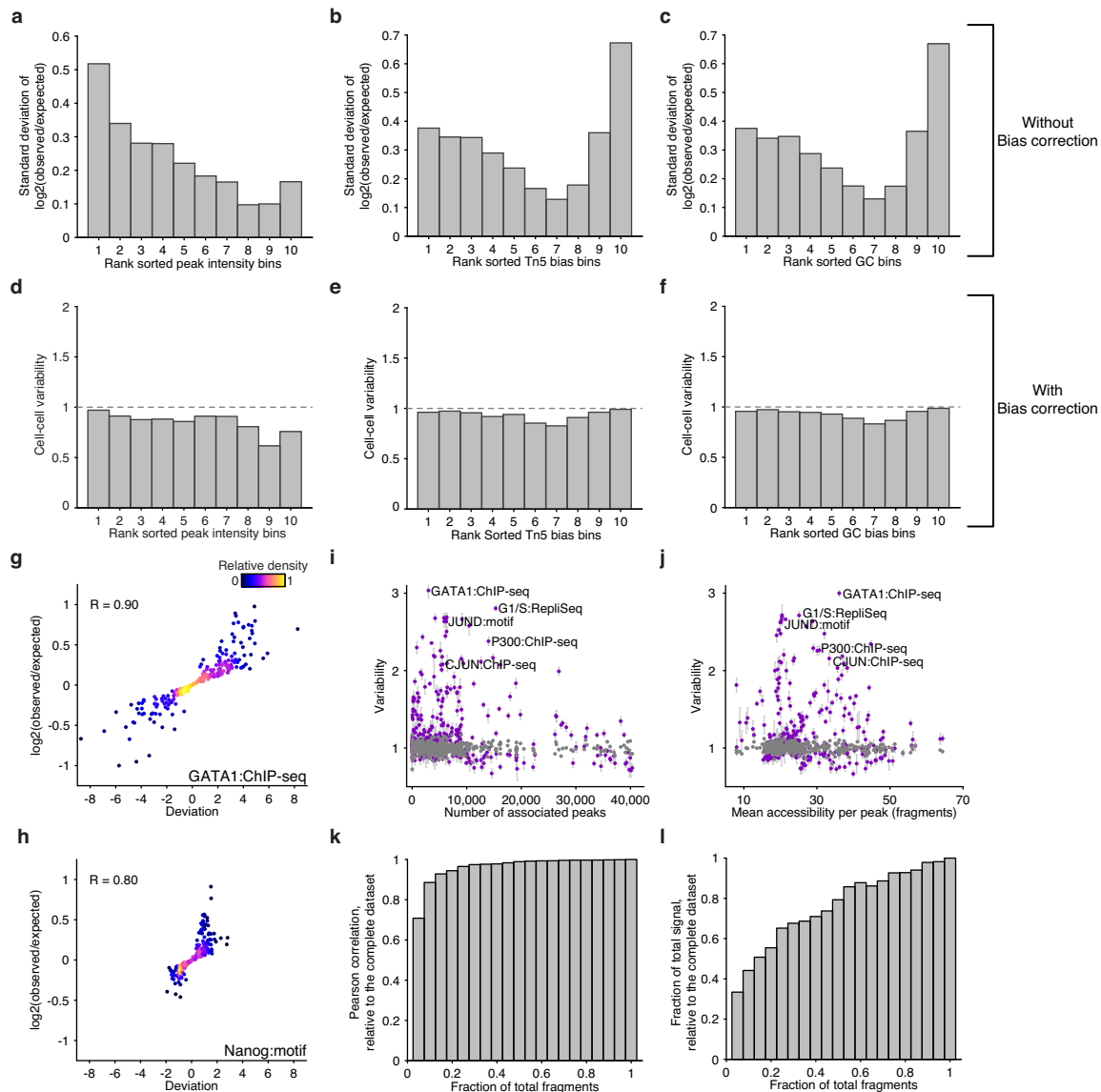
Extended Data Figure 2 | scATAC-seq data recapitulate bulk ATAC-seq characteristics. **a**, Fragments observed in open chromatin peaks identified from aggregate scATAC-seq data ($n = 384$ libraries) are highly correlated with reads observed from bulk ATAC-seq in GM12878 cells. **b**, Histogram of aggregated read starts around all transcription start sites (TSS) (in K562 cells) comparing ensemble approaches, including 500 cell ATAC-seq reported in a previous publication, to scATAC-seq shows high enrichment above background level of reads. **c**, DNA fragment size distribution of ATAC-seq fragments from single cells (grey) and the average of all single cells (red) display characteristic nucleosome-associated periodicity. **d**, Phase-contrast (left) and epifluorescence images (right) of captured cell no. 4 displaying characteristic

live cell stain (Calcein) and exclusion of ethidium bromide. **e**, Histogram of read starts around TSSs for cell no. 4 shows high enrichment. **f**, DNA fragment size distribution for cell no. 4 showing nucleosomal periodicity. **g**, Images similar to **d** showing staining of cell no. 83, suggesting low viability due to ethidium bromide staining. **h**, Histogram of read starts around transcription start sites shows lower enrichment than cell no. 4. **i**, DNA fragment size distribution for cell no. 83. **j**, Images similar to **d** showing staining of cell no. 33 suggesting viability. **k**, Histogram of read starts around transcription start sites of this cell shows low levels of enrichment. **l**, DNA fragment size distribution showing no nucleosome-associated periodicity.



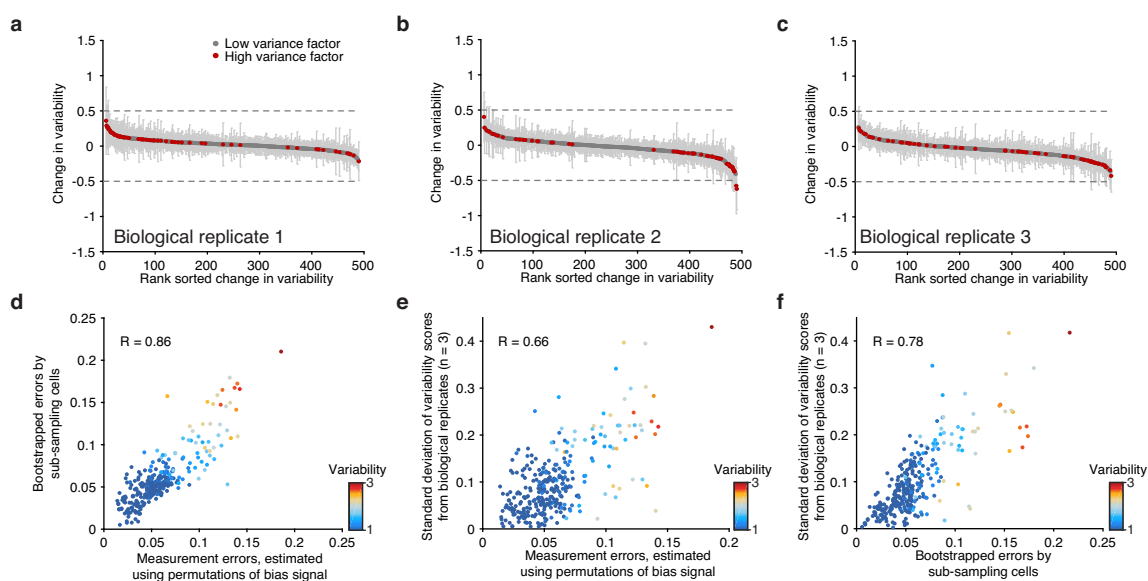
Extended Data Figure 3 | Fragment recovery metrics within scATAC-seq libraries. **a**, Accessibility across all peaks ($n = 50,000$) in GM12878 cells. **b**, Accessibility across all annotated promoters in GM12878 cells. Typical promoters used for subsequent analysis are boxed with dotted lines.

c, d, Recovery of typical promoters shown in **a** within single cells within observed (**c**) and extrapolated (**d**) data using measures of predicted library complexity.



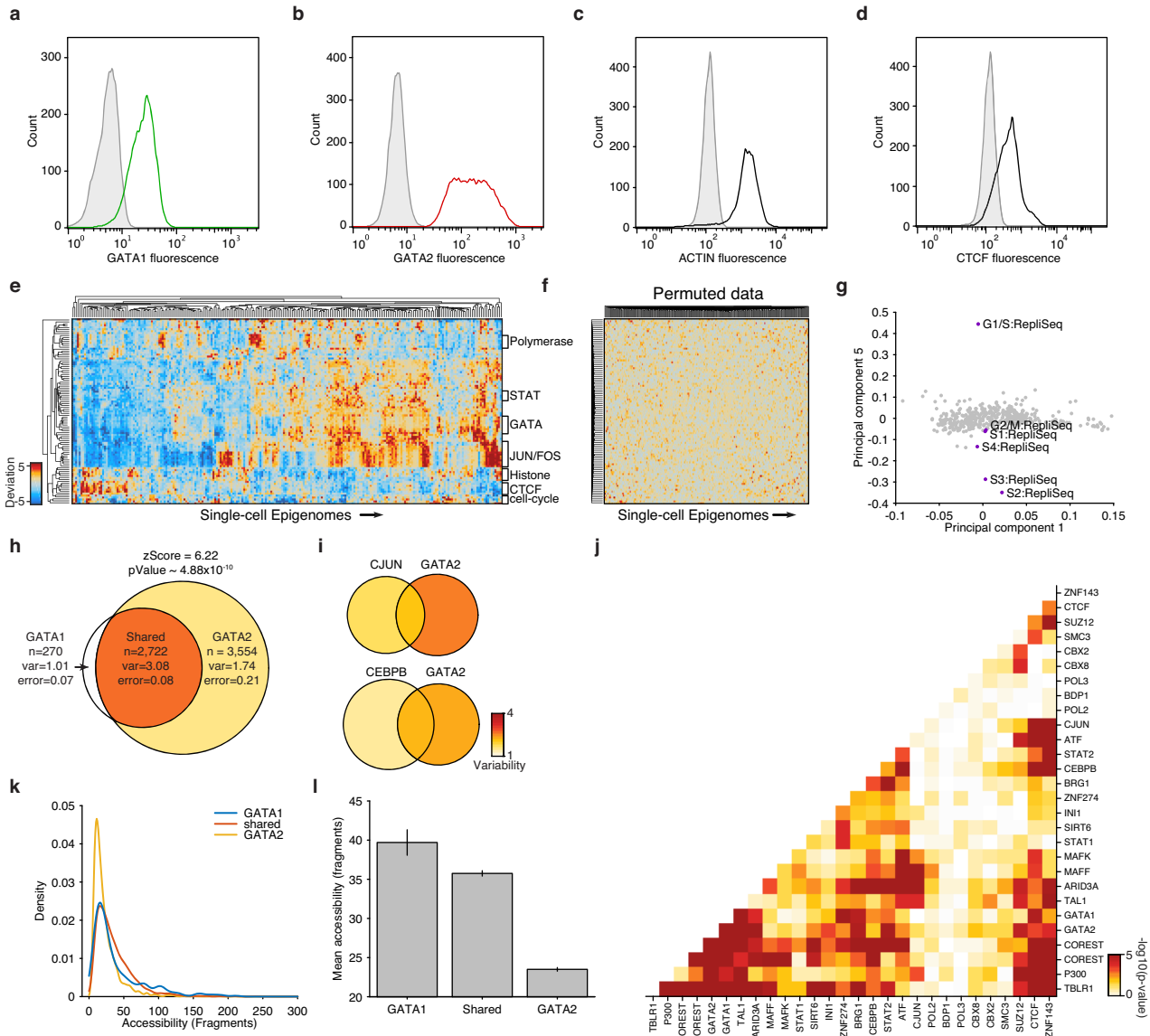
Extended Data Figure 4 | scATAC-seq data analysis pipeline and validation of bias normalization. **a–c**, Standard deviation of log-fold change in reads across cells within peaks binned by deciles of peak intensity (**a**), Tn5 bias (**b**) and GC bias (**c**). **d–f**, Variability scores (incorporating bias normalization) within the same peaks shown in **a–c**, peaks are binned by deciles of peak intensity (**d**), Tn5 bias (**e**) and GC bias (**f**). **g, h**, Log-fold change versus deviation scores across single K562 cells for GATA1 ChIP-seq target sites

(**g**) and peaks (**h**) containing a Nanog motif. **i, j**, Variability scores for factors (purple) and the permuted background (grey) ranked by number of peak associations (**i**) and the mean accessibility per annotated peak (**j**). **k, l**, K562 single-cell data sets showing the effect on variability scores as a function of downsampling fragments. Fidelity after downsampling is measured with correlation (**k**) and dynamic range (**l**) relative to the complete data set.



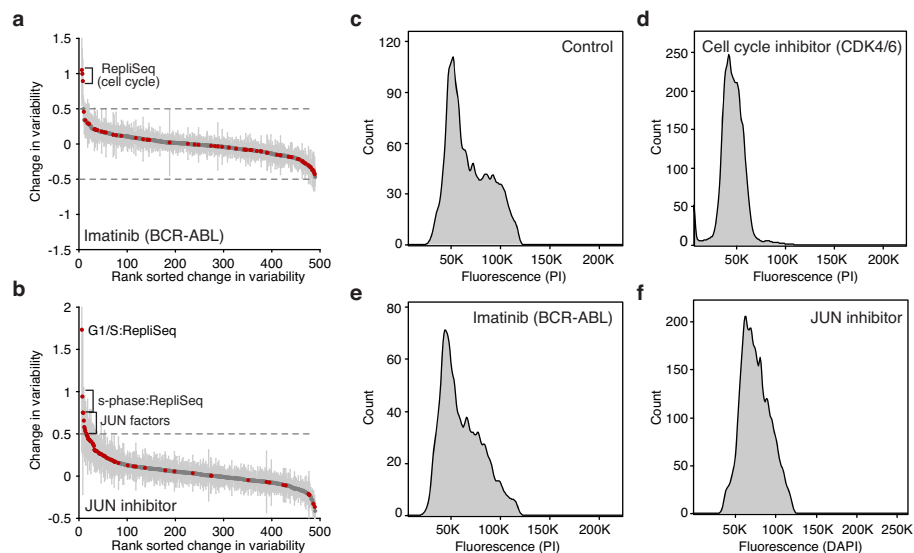
Extended Data Figure 5 | Biological replicates and measurement error analysis. **a–c,** Observed changes in variability comparing the merged set of replicates (K562) to each individual biological replicate. Error bars represent

one standard deviation of the variability scores after bootstrapping cells from each replicate. **d–f,** Correlation of errors computed using three distinct approaches.



Extended Data Figure 6 | Characterization of high-variance *trans*-factors in K562 cells. **a–d**, Distribution of GATA1 (**a**), GATA2 (**b**), actin (**c**) and CTCF (**d**) fluorescence observed by flow cytometry. Distributions in grey depict isotype controls. **e**, Bi-clustered heat map of single-cell deviations as observed within K562 cells ($n = 239$). Labels on right identify co-clustering of related factors. **f**, Bi-clustered heat map of single-cell deviations observed from permuted data. **g**, Projection of factor loadings onto principal component 1 versus 5 from principal component (PC) analysis of heatmap shown in **e**. Factor loadings do not vary along PC5, although peaks associated with regions with

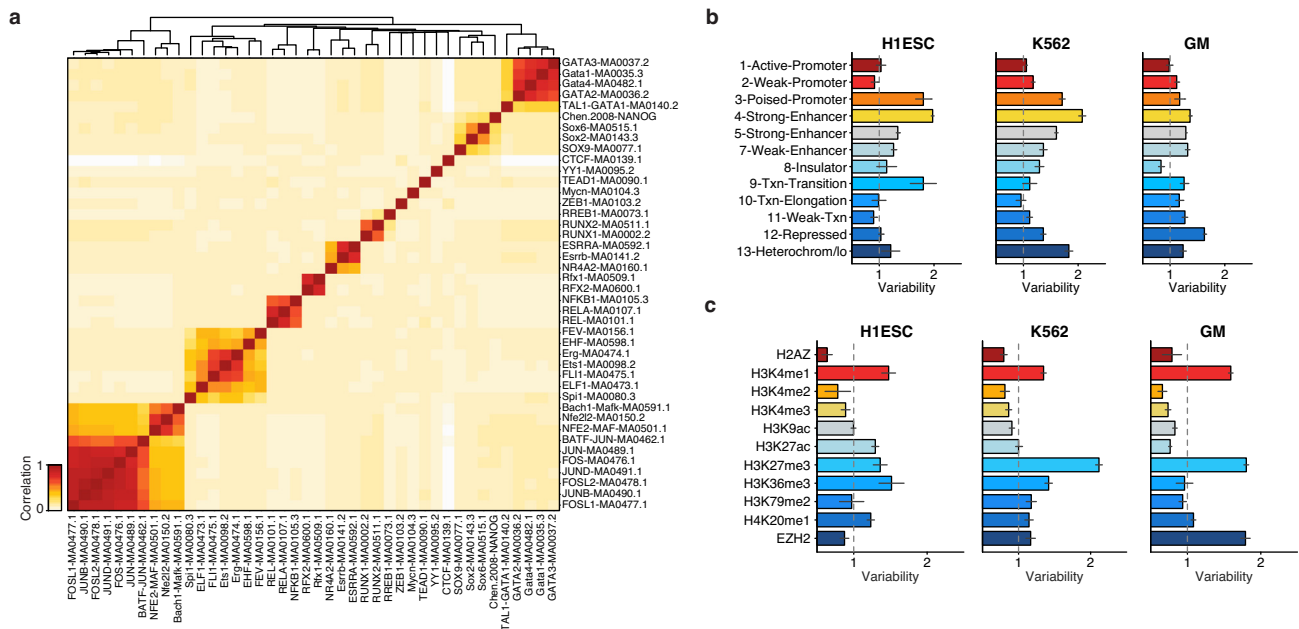
different replication timings (Repli-Seq) have strong variation along this axis. **h, i**, Venn diagrams showing variability of GATA1 and/or GATA2 (**h**), cJUN and/or GATA2 and CEBPB and/or GATA2 (co)-occurring ChIP-seq sites (**i**). **j**, The $-\log_{10}(P)$ values of calculated changes in co-occurring ChIP-seq sites shown in Fig. 2g. **k**, Distribution of accessibility among GATA1 only, GATA2 only, and shared sites. **l**, Mean accessibility from GATA1 only, GATA2 only, and shared sites in **k**, error bars represent one standard deviation generated by bootstrapping ChIP-seq peaks.



Extended Data Figure 7 | Drug treatments modulate factor variability.

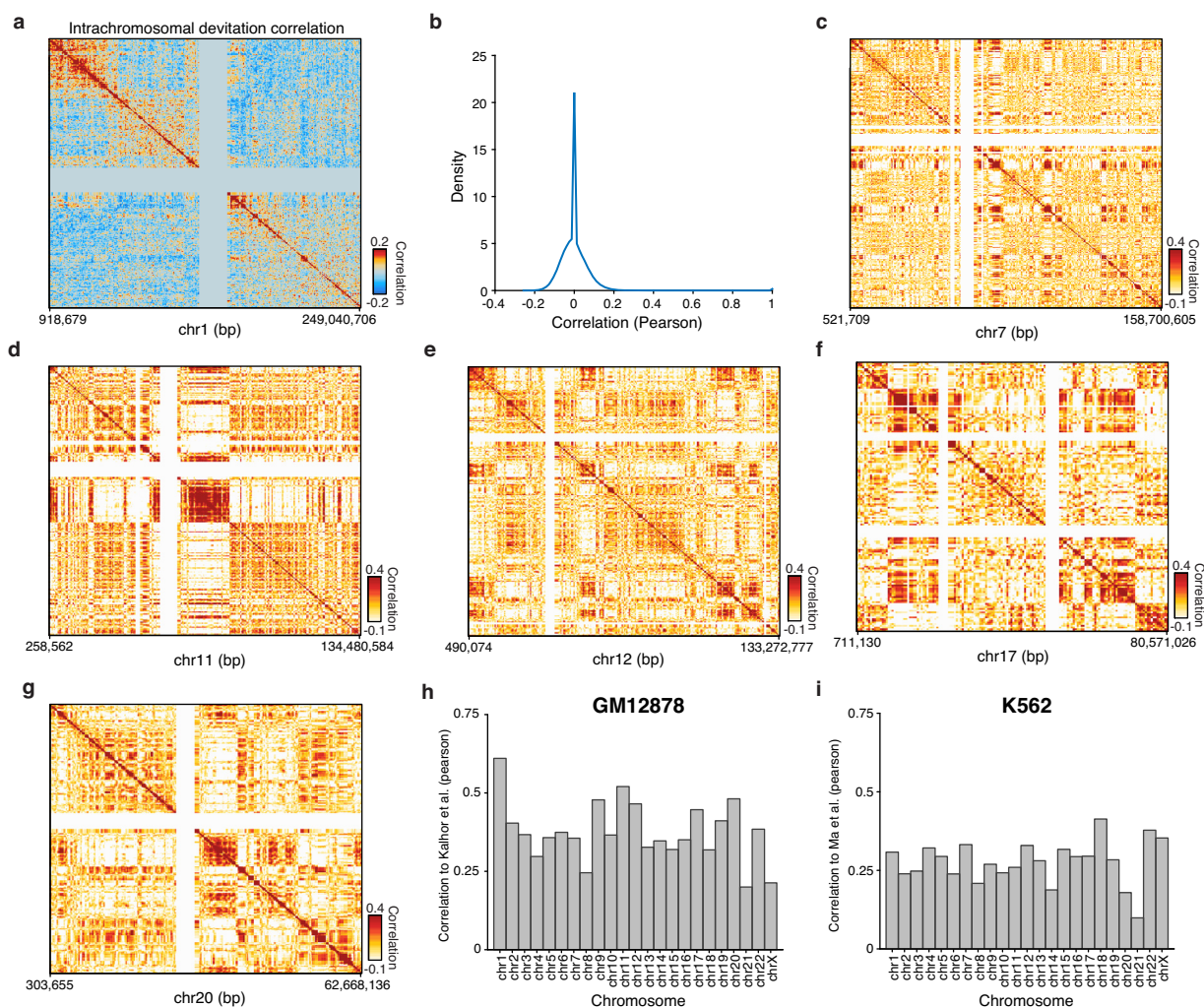
a, b, Change in variability of untreated K562 cells versus cells treated with imatinib (**a**) and JUN inhibitor (**b**) show increase of variability in factors associated with the cell cycle or S phase and JUN factors, respectively. **c–f,** Flow

cytometry data depicting DNA content, using DAPI or propidium iodide, in control K562 cells (**c**) or cells showing altered cell-cycle status after treatment with cell-cycle inhibitor (**d**), imatinib (**e**) or JUN inhibitor (**f**).



Extended Data Figure 8 | Transcription factor motif correlation and variability across chromatin state. **a**, Hierarchical bi-clustering of high-variance transcription factor motif annotations using the Pearson correlation.

b, c, Variability of regions associated with chromatin states (**b**), as identified in ref. 26, and histone modifications (**c**).

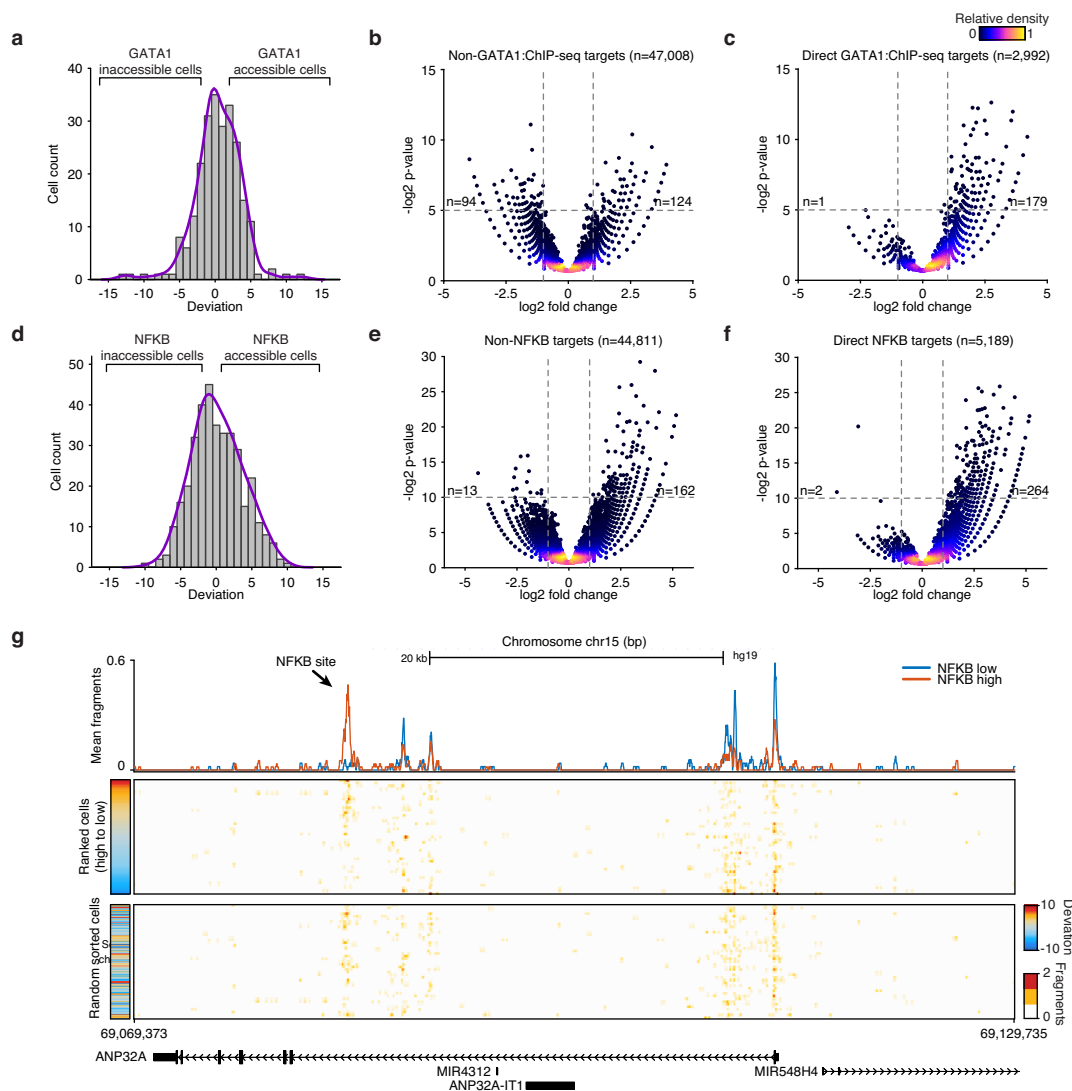


Extended Data Figure 9 | Cis-variability analysis within single cells.

a, Interchromosomal chromosome 1 co-correlations of deviation scores within single cells calculated for bins of 25 peaks within GM12878 cells.

b, Distribution, using density estimation, of correlation values shown in

a. **c–g**, Analysis of *cis*-correlation (identical to Fig. 4) for representative chromosomes 7, 11, 12, 17 and 20. Correlation between scATAC-seq *cis*-correlation and chromosome conformation capture methods for each chromosome in GM12878 (**h**) and K562 (**i**) cells.



Extended Data Figure 10 | Measurements of individual peaks within single cells. **a**, The distribution of GATA1 deviation scores for single K562 cells. **b**, **c**, Volcano plots of non-GATA1 (**b**) and GATA1 (**c**) peaks in K562 cells, P values were calculated using a binomial test. **d**, The distribution of NF-κB deviation scores for single GM12878 cells. **e**, **f**, Volcano plots of non-NF-κB (**e**) and NF-κB (**f**) peaks in GM12878 cells, P values were calculated using a

binomial test. Inset numbers show the number of points in upper left or upper right quadrants of the panel. **g**, Accessibility at a genomic locus, showing (top) aggregate NF-κB low (blue) and NF-κB high (red) profiles, (middle) single GM12878 cells ranked by NF-κB deviations scores and (bottom) unranked single cells.

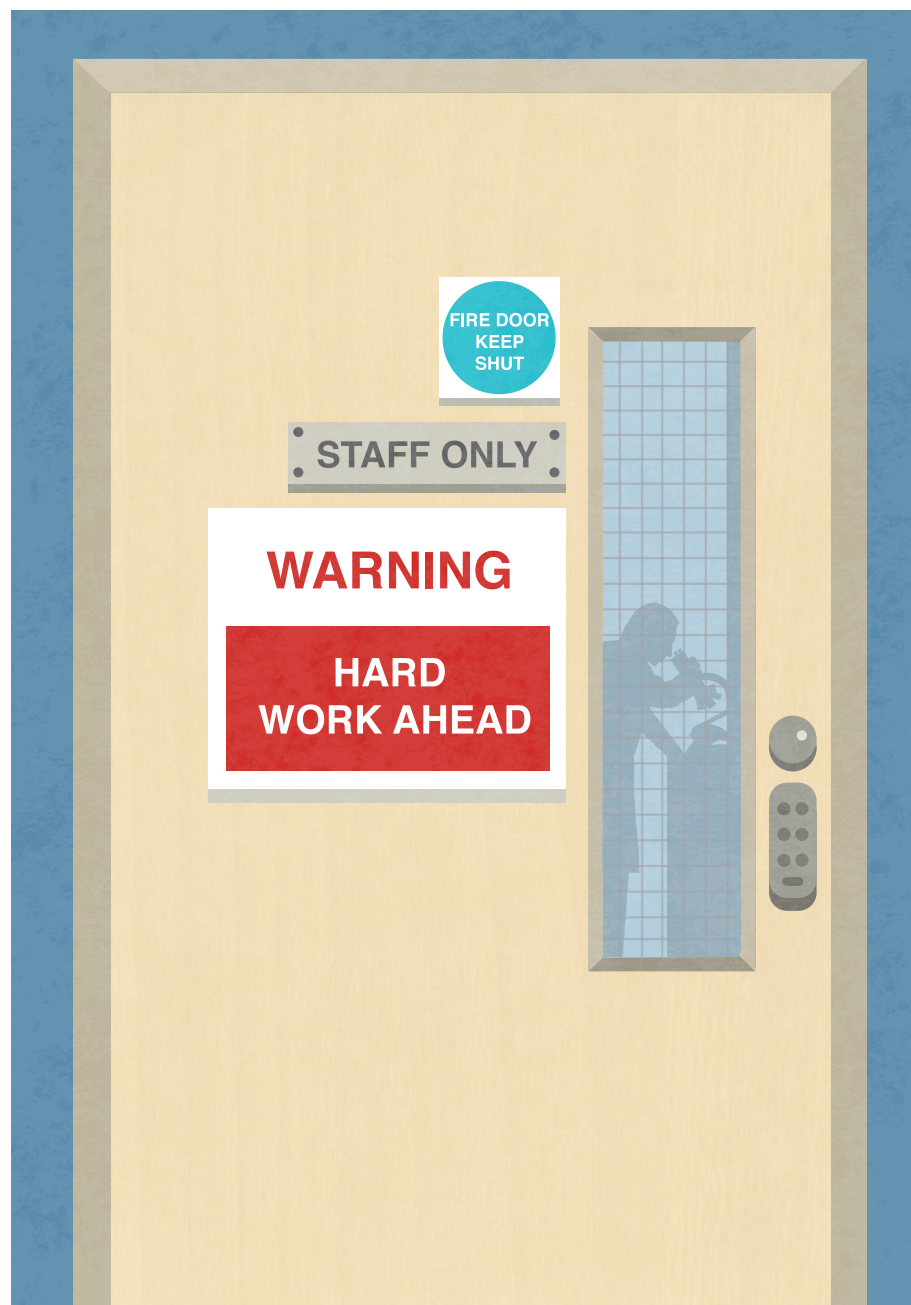
CAREERS

POLAR-BEAR CONSERVATION The role of science in a politically sensitive spot **p.493**

POSTDOC SURPRISE Another country, a new job and a funding freeze go.nature.com/e3lbuf

NATUREJOBS For the latest career listings and advice www.naturejobs.com

CLAIRE WELSH/NATURE



BY CHRIS WOOLSTON

When Andrew Hires looks back on his days as a graduate student and post-doctoral researcher, he wishes that somebody had told him how unpredictable science can be. “You do experiments, and 90% of them aren’t going to work. Nobody warned me about that,” says Hires, a neurobiologist at the University of Southern California in Los Angeles. He had to adjust his expectations, double down on perseverance and savour the successes when they came.

On the way from the first failed experiment to a tenured or tenure-track position at a major research institution, scientists who intend to remain in academia must learn about a lot more than just frogs, or photons, or whatever they are investigating — they must also accumulate hard-won lessons about publication, funding, promotions and a host of other subjects. Academic scientists who have already gone through the wringer have much to tell newcomers, but are likely to do so only if young scientists can put aside any ill-placed discomfort and ask. “They get plenty of guidance in the field and in the lab,” says Andrew Hendry, a newly tenured ecologist at McGill University in Montreal, Canada. But when it comes to the rest of the science life, he says, junior researchers are often stumbling in the dark, or at least walking slowly in a poorly lit room. “People don’t ask enough questions. They’re embarrassed,” he says.

It is crucial, experienced scientists say, that junior researchers ask questions of their mentors, supervisors, lab and department heads, senior colleagues and members of their network. The answers can add up to a handy guide for navigating up the ladder.

MAKE A MARK

Clara Nellist, a particle-physics postdoc at the Linear Accelerator Laboratory in Orsay, France, wishes that she had known how hard it is to stand out when working in large collaborations. And she would have liked to have known in advance how to turn all the meetings she is expected to attend into an advantage. Earlier this year, she co-authored an important paper that estimated the mass of the long-sought Higgs boson (G. Aad *et al.* *Phys. Rev. Lett.* **114**, 191803; 2015).

The problem: she had to share the glory with 5,153 other authors (see *Nature* <http://doi.org/4sn>; 2015). “The only person who’s ever going to find my name is my dad,” she says. ►

CAREER ADVANCEMENT

Insider knowledge

Junior researchers have a lot to learn, but talking to others about their experiences will help to avert nasty surprises.

► Still, she is making a name for herself in two projects: refining the pixel detectors at the Large Hadron Collider at CERN, Europe's particle-physics laboratory near Geneva, Switzerland, and studying the properties of the Higgs boson in ever-greater detail.

That combination of practical and theoretical physics should give her an edge in the job market, she says, although it doubles the number of meetings that she needs to attend. Many, many meetings — another surprise of the particle-physics world. “CERN knew that meetings were a problem, so they formed a committee to address it. The first thing the committee did was call a meeting,” she says.

Still, those meetings have given her a chance to add her voice to the field and to learn a few things in the process. At first, she was reluctant to speak up. “I would have questions, but I wouldn't ask them in front of an audience. I didn't want to admit to any gaps in my knowledge,” she says. An adviser finally pulled her aside to share one of the important life lessons of science: ask questions. “It shows



ANDREW HENDRY

Ecologist Andrew Hendry thinks that being a professor is the best job in the world.

that you're interested in the topic,” she says.

Chenjie Wang, a condensed-matter physics postdoc at the University of Chicago in Illinois, had even more trouble finding his voice. He managed to get his PhD without speaking to anyone other than his adviser. The language barrier was a problem, he says — Wang emigrated from China in 2007. But, more fundamentally, he had yet to appreciate how much other students would have to offer. “I had been told that Americans weren't very strong in maths and physics,” he says with a laugh.

He now sees that silence as a missed opportunity. Conversations around the labs at Chicago have given him a new outlook on science, and perhaps even a better understanding of maths. “Americans are wild thinkers, and they keep chasing answers,” he says. Where he might be content with a single solution, other researchers would continue to approach a problem from different angles, leading to new questions and possibilities. “It's very important to talk to people,” he says. “It will keep your mind open.”

Had he known that earlier, he says, he could have pushed his research — and himself — even further.

STRESS CONTROL

The value of conversation has also become clear to Christine Latin, a postdoc at Yale University in New Haven, Connecticut, who studies stress responses in live sparrows using radiological images. Her entire research project sprang from a chat about stress hormones that she had during a party.

She says that talking to other researchers has also helped to ease her loneliness, an aspect of the postdoc life that she wished she had been prepared for. “As a postdoc, you don't really have a cohort. You're on your own,” she says, adding that at least graduate students or their institutions regularly schedule social get-togethers.

It didn't help that her postdoc meant taking on a new project in a new lab, even a new city. “I didn't even know where the pipettes were,” she says. She soon found the pipettes, and, after a while, she found some like-minded people. She joined the board of Women in Science at Yale,

and reached out to faculty members, postdocs and graduate students who might be open to collaboration. “That has made me feel more like I am part of a research community,” she says.

Hoping to help junior researchers to avoid some common pitfalls and missteps, Hendry has written a 10-part (and counting) series of posts on his popular Eco-Evo Eco-Evo blog (see ‘How to get ahead’). He drew on personal experience for a post on how to choose a journal. At the time, he had been involved in 45 manuscripts that were submitted to top-tier journals, and had just one accepted. As he notes in his blog, “rejection is an ever present companion in science.”

Too many proposals and papers lack a sense of purpose, he says. His advice: every piece of scientific writing should employ the ‘baby-werewolf-silver-bullet formula’. In other words, the work should have a clear problem (the werewolf), a definitive solution (the silver bullet) and a strong sense of the stakes (the baby). “There has to be something we all care about,” he says.

Hires wishes that he had known more about when to publish, not just what. As a graduate student, he created a sensor that measures the release of glutamate in brain cells. That development was certainly worthy of a paper, but he decided to wait until he could demonstrate the sensor in a key experiment, a process that took another four years.

In the meantime, someone else came up with the same idea and published a paper before Hires ever got his experiment to work. “I got scooped,” he says. “If I had been more realistic and less ambitious, I would have published immediately and left the application for another paper.”

Andrew Jackson, a theoretical ecologist at Trinity College Dublin, wishes that he could have given his younger self advice about writing his first grant proposal. “I was naive. I thought, here's a cool idea that I'll just throw at you. I fired it off, and it came back covered with comments. I was rightly slammed for it.” The lesson: “Have senior colleagues read your proposal.”

He recovered from that misstep and was in time offered a tenured position. But he had yet to learn the art of negotiation. “When I got

ADVICE

How to get ahead

Junior researchers tend to have many questions about science — so many that they do not always know what to ask. Andrew Hendry, an ecologist at McGill University in Montreal, has tried to fill in those knowledge gaps with a series of ‘how to’ posts at his popular Eco-Evo Eco-Evo blog (ecoevoevoeco.blogspot.ca). Topics include ‘how to do statistics’ and ‘how to respond to reviewers’. Here are some highlights.

- Don't throw out your data just because they don't seem to fit a particular statistical model. “The data are the real thing,” he says. “The stats are just a tool to aid in interpretation.”
- For maximum citations, he advises, don't be afraid to submit a good study to a top-tier journal, even if rejection seems likely. The challenge will inspire you to make the paper as solid as possible, and the reviews might sharpen it even more. This route does entail more time and stress, but the potential payoff is greater.
- Beware the ‘grass-is-greener’ syndrome. It can be tempting to give up a messy, complicated project for a venture that seems clear-cut and straightforward. But the new project is bound to have problems, too. “A given project always looks best before the actual work starts,” he says.
- Whenever possible, he says, finish what you start, and publish what you finish. **C.W.**

the call, they asked me what kind of salary I wanted. I suggested a number, and they immediately accepted," he says. "I realized later that if people aren't saying 'no' to you, you aren't asking for enough."

SELF CONFIDENCE

Meghan Duffy, an evolutionary biologist at the University of Michigan in Ann Arbor, is a self-assured, widely acclaimed authority on aquatic ecology. But as a graduate student, she had many of the same doubts that plague other young researchers. Was she really cut out for this business? How could she live up to the standards of the senior scientists around her? And, most pressingly, would she ever catch up on the maths?

She started her graduate work with just one university maths course and no programming under her belt. "My undergraduate self didn't realize that those skills would be useful," she says. Duffy checked out a book on calculus, did a theoretical-ecology course to get a better grasp of the mathematical side of aquatic science and taught herself how to program.

More importantly, however, she learned that she was not the only one to doubt herself — and that there was usually no need. "So many people have impostor syndrome," she says. "You can't compare the turmoil inside you to someone else's confident exterior."

Hendry, too, advises an upbeat outlook. "Being a professor is probably the best job in existence," he says. "The research, the day-to-day life, it's all up to you. I can't imagine a job with more freedom than I have."

Groundbreaking research at the top of a field is a hyper-competitive arena, he says, but a scientist can do great work without a huge level of stress. "I'm not a global high-roller," he says. "You can have a more relaxed and fun life."

He now also has a take on science that would have come as a surprise to his younger self — and a lot of other junior researchers. "If you really want to be a professor, and you have a half-decent research record, and you aren't picky about where you want to work, you will eventually get a job," he says. "Don't give up."

So getting ahead in science is easier than many people think. Junior researchers often try to work things out for themselves, but if they seek out advice, they will find that people are willing and eager to share what it takes to succeed. They just need to ask. ■

Chris Woolston is a freelance writer in Billings, Montana.

TURNING POINT Mike Runge

US Geological Survey (USGS) wildlife ecologist Mike Runge co-chairs a team that released a species-recovery plan on 6 July for the polar bear — one of the first high-profile mammals to be listed as threatened in connection with climate-change projections. He explains how he learned to balance science with policy.

What best prepared you to work in science policy?

I taught secondary school for five years after getting my undergraduate degree in molecular biology and philosophy. You can't teach calculus to 17-year-olds at 8 a.m. unless you think about their motivations and what will engage them. It was an extraordinarily valuable job that taught me how to listen, be fair-minded and communicate effectively with people in different settings — skills that are crucial for me today.

How did you get into the field?

In my PhD programme in wildlife science, my project was to develop quantitative models of beaver-population dynamics. I combined population models with factors that affect trapping efforts, such as pelt price and cost of gas, which New York wildlife managers could use to help to regulate beaver trapping. I use that approach to integrate quantitative scientific methods into real-world settings.

When did you begin to work on polar bears?

After I started a postdoc with the USGS, I got a call to work on predictive population models of manatees. They are protected under both the US Endangered Species Act and the US Marine Mammal Protection Act, so I learned about the legal frameworks under which science is used. In 2007, when the US Fish and Wildlife Service was petitioned to list polar bears as threatened, the USGS was tapped to study the bears' current and projected population numbers in the face of climate change. Thanks to my manatee experience, I was asked to join the polar-bear population-modelling team and, later, the recovery-plan team.

What was the biggest challenge in writing the recovery plan?

Polar bears are an icon of the Arctic. Everybody cares about them. And there are diverse groups of people — from Alaska's natives to its oil-and-gas industry to Polar Bears International, an advocacy organization — that are passionate about different aspects of the Arctic. It was a challenge to identify and bring together such a broad array of voices so that each one can be heard. But it is also an opportunity to create



something enduring — a shared vision for how polar bears should be managed.

Were negotiations tense?

The hard part was getting everyone to lay their cards on the table so that we could work out a solution together. I've tried to create a template for facing contentious issues and trade-offs straight on and with respect. At first, I thought that the politicians at the top would be the worst when it came to cooperation. But I've found that many lawmakers — those who work on the big, highly visible issues every day — know how to disagree respectfully and seek solutions through compromise.

What advice would you give to scientists who work in politically sensitive areas?

To be fair-minded. I've had the privilege to meet with different groups and to learn how aspects of wildlife management affect their lives. Lots of these situations are political and tense and require a non-judgemental understanding of multifaceted interests in the natural world.

Are you writing the rules for how climate-change-related recovery plans should proceed?

We didn't set out to create the template. The US National Oceanic and Atmospheric Administration fisheries service released a recovery plan for coral species early this year, which was the first to address species-level impacts and mitigation strategies in the face of climate change. We want to do the best that we can for polar bears. I think that a number of other recovery teams will have to go through a similar process. ■

INTERVIEW BY VIRGINIA GEWIN

This interview has been edited for length and clarity.

THE MEMORY OF TREES

An unnatural request.

BY LYNETTE MEJÍA

The old man held the brush above the paper, his fingers trembling as the sound of footsteps grew louder in the hallway. Around him the pale walls shivered, their pastel colours twisting and swirling together in rapid succession.

A guard entered, striding with broad, confident steps over to where the wizened figure sat in front of the easel. He glanced down at the thick, creamy paper affixed to its surface, his eyes momentarily flicking to the walls. Then he looked at the old man and sniffed, wrinkling his nose as if detecting a slightly offensive odour.

"Nothing yet, I see," he said.

"No," answered the old man.

The guard crossed his arms. "I don't understand what the problem is," he said, beginning to pace. "We explained it all to you. Anything you can visualize, anything at all, appears on the walls around you. Just paint what you see."

"Yes, I know," said the old man. He rubbed his forehead with one hand, felt one of the implants bulging slightly beneath the skin. "I tried to explain to the young man who..." He looked up at the guard and swallowed. "...interviewed me when I arrived here. That's not how it works, you see. That was never how it worked."

The guard took a deep breath and crossed the room, settling himself upon the small, hard cot in the corner. Reaching into a pocket he pulled out a packet of cigarettes and a lighter, sighing with pleasure as he lit one and pulled the smoke into his lungs.

"You don't mind?" he asked, although the old man knew he wasn't really asking. "Synthetic, of course, but I figure you might appreciate it. A little taste of home, eh?" He smiled, tapping grey ash onto the floor. "That's how I got this assignment, you know.

It's because I like old-fashioned stuff, collecting things. Always been a hobby of mine."

He smiled again,



although his eyes were distinctly colder now. Leaning forward, he tossed the cigarette onto the floor before grinding it under his heel.

"See, the thing is, Mr Bradstreet, you've been here a long time now. People are starting to lose patience. I'd say you need to produce something soon. Do you understand what I'm saying to you? You need to make a painting, a drawing, something."

"I can do an abstract," the old man said wearily, "but not landscapes. Not in here."

"Oh, I beg to differ," the guard said, rising. "We don't need abstracts, Mr Bradstreet. We did our research. You were quite the renowned landscape artist in your day. It's why we spent all that money to bring you back. You were nothing but a pile of bones, on the fast track to oblivion. Now you have a chance to be remembered for ever."

The artist looked down at the floor. He felt so tired.

"Cloning has its issues," the guard said, as if reading his thoughts. "Life spans are considerably shorter. If you don't give us something, we'll have to start this whole goddamned process over again, and I can promise you, the next go-round won't be nearly as pleasant as this one was."

He came closer, and the old man cringed as he leaned in.

"We got you the paints," he said in a low, snarling voice. "We got you the brushes. Do you know how hard it was figuring out the ingredients? We had to take paint samples from the few surviving fragments, do full spectral analysis, reproduce the compounds. Suffice it to say that the whole thing was very expensive."

The old man stared blankly at the wall in front of him. "I need to go outside," he said finally, turning. Around them the walls suddenly danced with colour and shadow, the images resolving themselves into a lush green landscape full of trees and flowers. Overhead the sun shone in a cloudless blue sky. Flowers nodded in a phantom breeze, while bees buzzed lazily between nodding blossoms.

The guard clapped. "There you go! That's perfect! Now just paint that!"

The old man looked up at him, his face sad and tired. "I need to go outside," he repeated. "This isn't the same, don't you see? I can't paint the memory of trees."

The guard sighed, walking slowly to the door. "We're dying, Mr Bradstreet," he said. "What's left of us is committing suicide by the thousands every day." He ran a hand through his hair. "We have the tech, you see, but we've lost the ability. Photos just aren't cutting it. We need to feel it. Can you understand that now that you've been here a while?" From his pocket he pulled a small device and pressed a button on its side. Instantly one of the walls went transparent, revealing a burned and blackened landscape populated only by sparse patches of dried, dead grasses. The old man began to weep softly.

"You see Mr Bradstreet?" the guard said as he turned the knob to go. "You are the memory of trees." ■

Lynette Mejía writes science fiction, fantasy and horror prose and poetry. Her work has been nominated for the Rhysling Award and the Million Writers Award. You can find her online at www.lynettemejia.com.

ILLUSTRATION BY JACEY